November 6, 2020 BLA 761178: Peripheral and Central Nervous System Drugs Advisory Committee Meeting

Script for FDA Presentation: Aducanumab for the Treatment of Alzheimer's Disease: Clinical Overview of Efficacy

FDA Presenter: Kevin M. Krudys, PhD

Slide 1

Hello, my name is Kevin Krudys and I am a clinical analyst in the Division of Neurology 1. In this presentation I will provide a clinical overview of the evidence provided to support the effectiveness of aducanumab for the treatment of Alzheimer's disease.

Slide 2

Aducanumab is an anti-amyloid beta monoclonal antibody targeting aggregated forms of amyloid beta. Accumulation of amyloid beta in the brain has been proposed to be the primary driver of Alzheimer's disease and has been widely considered a target for drug development. Anti-amyloid therapies have thus far had a disappointing record with many high-profile, late-stage failures. It is important to keep in mind, however, that anti-amyloid therapies are not a distinct class of drugs but rather can have many different characteristics. Therefore, previous failures are not particularly informative for the assessment of the effectiveness of aducanumab. Aducanumab is distinguished in some ways by its molecular characteristics and by features of its clinical development. Aducanumab's origin, based upon reactivity to aggregated amyloid beta in elderly subjects who appear phenotypically resistant to cognitive deterioration is unique amongst agents targeting amyloid beta. Throughout development, aducanumab clinical trials have enrolled patients at earlier stages of the disease with evidence of amyloid-beta pathology. Prior to initiation of large efficacy and safety trials, an early study demonstrated target engagement and confirmed reduction of amyloid-beta plaque burden.

Slide 3

Three clinical studies are relevant for the evaluation of efficacy of aducanumab and will be discussed in further detail in this presentation. Studies 301 and 302 were large, global, double-blind, placebo-controlled studies of identical design which were intended to evaluate the efficacy and safety of aducanumab. Study 103 was a smaller, dose-ranging clinical trial conducted in the United States. Although designed primarily as a safety and tolerability study, Study 103 was an adequate and well-controlled study that explicitly included assessment of prespecified clinical and biomarker endpoints and a prospectively defined statistical analysis plan.

Slide 4

The table on this slide summarizes a few of the key elements of Studies 301 and 302. A detailed description of the protocol design is provided in the briefing document. These studies enrolled patients with mild cognitive impairment due to Alzheimer's disease and mild Alzheimer's disease dementia with a positive amyloid PET scan at baseline. These diagnostic criteria are consistent with Stage 3 and Stage 4 as defined in the 2018 FDA draft guidance. The placebo-controlled period was 78 weeks in duration and two dose levels (which I will refer to as low and high) were evaluated. Dosing was based on the patient's apolipoprotein or ApoE e4 carrier status. The primary endpoint was the change from baseline in the Clinical Dementia Rating-Sum of Boxes (or CDR-SB) score at Week 78. CDR-SB is a scale that adequately and meaningfully assesses both daily function and cognitive effects in an integrated manner and is consistent with FDA guidance on clinical endpoints appropriate for this population. The Division accepts a statistically significant change on an inherently meaningful instrument such as CDR-SB as evidence of a clinically meaningful effect. Secondary endpoints include the Mini-Mental State Examination (or MMSE), the Alzheimer's Disease Assessment Scale – Cognitive 13-Item Scale (or ADAS-Cog 13) and the Alzheimer's Disease Cooperative Study – Activities of Daily Living-Mild Cognitive Impairment (or ADCS-ADL-MCI). The protocols were granted Special Protocol Assessment (or SPA) agreements in 2015. SPA agreement indicates concurrence by FDA with the adequacy and acceptability of specific critical elements of overall protocol design such as entry criteria, dose selection, endpoints and planned analyses.

Slide 5

Before discussing the results of these studies, it is critical to address the public declaration of futility and subsequent termination of the studies in March of 2019. The dataset that was used to inform the futility analysis included data from all patients enrolled in the studies who had the opportunity to complete the study by December 26, 2018. The futility criteria were based on conditional power, which was defined as the chance that the primary efficacy endpoint analysis would be statistically significant in favor of aducanumab at the completion of the studies, given the data at the interim analysis. Conditional power was calculated with two assumptions. First, that the treatment effect would be similar in the two studies and second, that the treatment effect would not change over time. The studies were to be considered futile if both studies had conditional power less than 20% for both dose groups. The analysis demonstrated that the futility criteria were met, and the applicant followed the prespecified plan by announcing the termination of the studies on March 21, 2019. It is important to note that declaration of futility does not mean that the studies had failed. It simply means that based on certain assumptions, it was unlikely for the studies to be successful if run to completion.

Slide 6

In fact, the futility decision was not an accurate reflection of the individual studies. Calculation of conditional power pools the observed interim data from the two studies to predict the future unobserved effect. Central to this calculation is the assumption that the treatment effect would be similar in the two studies. This is a reasonable assumption, as the studies had identical protocols and were run concurrently. And in fact, this assumption was upheld in the low dose treatment arms as seen in the table. The treatment effects for the low dose arms were -0.15 and -0.10 for Study 301 and 302,

respectively. In this table, negative numbers reflect a treatment effect in favor of aducanumab and positive numbers reflect a treatment effect in favor of placebo. The treatment effect for the high dose arms were clearly divergent, with a treatment effect of 0.22 in favor of placebo for Study 301 and a treatment effect of -0.28 in favor of aducanumab in Study 302. In other words, the assumption that the treatment effect would be similar in the high dose arms of the two studies was violated.

Slide 7

Upon being informed of these results in May of 2019, the Division asked the applicant to recalculate conditional power for each study independently, without pooling data from the two studies. The results are provided in the table and show that the futility criteria would not have been met with this analysis. In fact, this analysis suggests that Study 302 would have had a reasonable chance of success if run to completion as indicated by a conditional power of 59% in the shaded blue cell of the table.

Slide 8

It is important to note that between the data cutoff in December 2018 and the futility announcement in March 2019, rigorous, per-protocol collection of blinded data continued in the studies as planned. Subsequently, the primary analysis per the statistical analysis plan was conducted on this more complete ITT dataset including all assessments collected before March 21, 2019. On face, Study 302 was a positive study on several of the instruments used to evaluate efficacy. Study 301 was a negative study. To a casual observer, the turnabout from futility to positive study might suggest that the data collected after the futility announcement was completely divergent from the data used to make the futility decision or that some new type of analysis was performed on the data. The high-dose treatment effect in Study 302 did improve over time (from -18% with the futility dataset to -22% with the final dataset) but one can see that this study was always trending positively. The magnitude of the effect in the high-dose arm in Study 301 underwent the greatest change from the analysis with the futility dataset to the analysis with the final dataset, 15% to 2%, suggesting instability in the estimate. Treatment effect clearly improved over time and especially in Study 301.

Slide 9

The first formal interaction between the Applicant and FDA regarding the termination of the studies and subsequent analyses was in the form of a Type C Meeting in June of 2019. In their meeting request, the Applicant simply asked the Division for advice on next steps. Upon review of the data, the Division noted that it would have been more appropriate if futility had not been declared and that it was possible that Study 302 might not only be interpreted as being supportive of the efficacy of aducanumab but might also be considered exceptionally persuasive. Given the unmet medical need and the possibility that aducanumab is an effective drug for the treatment of Alzheimer's disease, the Division advised that extensive resources should be brought to bear in a collaborative manner to achieve a maximum understanding of the existing data. A critical first step before further consideration could be given to

Studies 301 and 302 was to evaluate the effect of early termination of the studies on the interpretability of the observed efficacy data and associated analyses.

Slide 10

To address this critical first step, virtual completion of the studies via modeling and simulation was used to explore the range of plausible outcomes had the studies been run to completion. Two approaches were used to virtually complete the studies. Both approaches estimated parameters from a target dataset, simulated patient-level data and then fit a model to the data. The primary approach supplemented the existing observed data with simulated assessments for the data that were censored due to the early termination of the studies. Another approach fully simulated studies to explore the range of plausible results if many studies like 301 and 302 were run from start to completion. Overall, simulation results were highly consistent with the primary analysis of the observed data. The results of this exercise established that the early termination of the aducanumab program did not compromise the interpretability of the efficacy results for Studies 301 and 302. The appropriate dataset for further consideration was agreed to be the ITT population consisting of all randomized subjects who received at least one dose of study treatment and excluding data collected after March 20, 2019. These conclusions were formalized in a Type C Meeting in October of 2019.

Now, I will present the results of Study 302 and 301. These results simply provide the final, interpretable results of studies that were terminated early but analyzed in accordance with the prespecified analysis plan.

Slide 11

Study 302 was a positive study and met its primary and secondary objectives. This table summarizes the placebo decline and the week 78 differences versus placebo for the primary endpoint (CDR-SB) and the three secondary endpoints. Again, negative numbers signify a treatment effect favoring aducanumab. The high-dose treatment arm demonstrated a statistically significant treatment effect compared to placebo of 0.39 points – which corresponds to a 22% difference - with a p-value of 0.012. This treatment effect was similar to the estimate used to power the study, 25% less decline. Statistically significant treatment effects in favor of aducanumab high dose were also observed for all three multiplicity controlled secondary endpoints. Positive results were also observed for the Neuropsychiatric Inventory-10, or NPI-10, the tertiary clinical efficacy endpoint which assesses the presence, frequency and severity across different neuropsychiatric domains. Numerically favorable results were observed for three of the four clinical endpoints in the low dose arm. The estimate of the treatment effect for CDR-SB in the low dose arm was -0.26, or -15%, although the p-value, 0.09, did not quite reach statistical significance. In general, the results of the low-dose arm are consistent with a dose-response relationship.

These results were robust to a number of prespecified and post hoc sensitivity analyses, including one which only included patients who had the opportunity to complete the study before its termination.

Slide 12

Study 302 actually encompasses two acceptable approaches to establish effectiveness which are described in the 2018 FDA draft guidance for developing drugs in early Alzheimer's disease. First, FDA accepts a statistically significant change on an inherently meaningful instrument such as CDR-SB as evidence of a clinically meaningful effect. This is highlighted in the green shaded cell in the table. Further clinical support is found in individual domains of CDR-SB which were consistent with the overall result. Second, FDA accepts an approach that independently assess cognition and daily function. ADAS-Cog13 and ADCS-ADL-MCI have been used as co-primary endpoints in Alzheimer's disease studies. It is worth noting that secondary endpoints were assessed by a second rater who was independent from the rater responsible for administering the CDR-SB. Although there is some overlap between the 4 clinical efficacy endpoints, each also captures distinct information regarding cognitive decline.

Slide 13

The figure on this slide illustrates the longitudinal change from baseline in CDR-SB in Study 302 for the three treatment arms. The effect on CDR-SB did not clearly emerge until the Week 78 visit with favorable trends at earlier time points, but not clear separation.

Slide 14

Although the precise relationship between reduction in brain amyloid plaque and clinical outcomes in unknown, consideration of the time course of reduction in brain amyloid (as measured by PET SUVR) is instructive. The plot on this slide illustrates the longitudinal change in PET SUVR for a typical patient receiving the three dosing regimens that were included in Study 302. The vertical axis represents the brain amyloid burden as measured by PET SUVR and the horizontal axis represents the time in the study in weeks. The solid black vertical line at 78 weeks represents the end of the placebo-controlled period. The red dashed horizontal line is a value for SUVR reported to discriminate between an amyloid positive and negative scan. The three solid lines represent profiles for a typical patient receiving target doses of 3 mg/kg, 6 mg/kg, or 10 mg/kg after titration as described in the protocol. The first key observation is that the reduction in brain amyloid is delayed with respect to initiation of treatment and it is similar in the three dosing regimens as seen in the first 26 weeks of the study. Delays are introduced by titration, accumulation of aducanumab exposure to steady state levels and a delay between concentration and effect as measured by SUVR. Titration was used to minimize the occurrence of ARIA, but at the expense of a delay in plaque reduction. Therefore, it is not surprising if a clear treatment effect was not observed at early time points. The second key observation is that the effect of aducanumab on amyloid plaque is increasing throughout the study and does not reach a maximum until after the placebo-controlled period. This suggests that the treatment effect should be durable and increase over time, but also raises the question of whether the placebo-controlled period should have been longer to more adequately assess efficacy. The figure does not take into account further downstream effects that may be attributable to effects on downstream markers of tau accumulation. It is also worth noting that this figure shows a typical individual for the purposes of illustration. Within a population there will be variability due to other factors, including dose modification due to ARIA.

Slide 15

Overall, the results of Study 302 are highly persuasive and capable of providing the primary contribution to a demonstration of substantial evidence of effectiveness of aducanumab. Study 302 is a strongly positive study on multiple distinct and important clinical measures and robust to numerous sensitivity analyses.

The treatment effect in Study 302 is supported by consistently favorable results for the primary and secondary endpoints across 79 of 80 prespecified subgroups of interest defined by demographic and baseline disease characteristics. Brain amyloid beta measured by PET was significantly reduced in a dose-dependent manner in all prespecified subgroups. Biomarkers reflecting target engagement, effects on downstream Alzheimer's tau pathophysiology (such as phosphorylated-Tau and Tau PET), and neurodegeneration (as measured by total-Tau) supported the observations on the clinical outcomes. Dose- and concentration-dependent relationships for biomarkers offer further support to the apparent dose- and concentration-response relationships for clinical endpoints. Detailed information for these findings can be found in the briefing book.

Slide 16

Study 301, on the other hand, was a negative study. The primary efficacy endpoint analysis, change from baseline in CDR-SB at Week 78, did not show a difference in the aducanumab high-dose treatment arm compared to placebo. The estimate of the treatment effect was 0.03, or 2% in favor of placebo over aducanumab. A numerical difference of -0.18 or -12% in favor of the aducanumab low-dose treatment arm compared to placebo was observed for the primary endpoint. Numerically favorable results for the low dose were also observed for all secondary endpoints. The effect on biomarkers for the high-dose arm in Study 301 was smaller than that observed in Study 302, but the difference was not large enough to explain the differences in the clinical outcomes.

Slide 17

The table in this slide shows the results of Studies 301 and 302 side-by-side. One can see that placebo decline for CDR-SB in Study 302, 1.74 points, was greater than in Study 301, 1.56 points. But differences in placebo decline do not appear to explain the difference in the outcomes of the studies. Placebo decline was actually lower for MMSE in Study 302 compared to Study 301 and the placebo decline was similar for ADAS-Cog 13 between the studies. Consistently positive outcomes were observed across the endpoints in Study 302, not just ones with a greater placebo decline.

Upon initial review, the one positive study (Study 302) and one negative study (Study 301) were given equal weight and consideration. Faced with these circumstances, a reasonable reaction is to desire an additional study with the expectation that it will provide a more definitive conclusion. If the results of Study 302 had not been robust, or if results on the secondary endpoints for Study 302 were inconsistent, or if biomarkers did not suggest an effect on disease pathology, or if the low dose treatment effect estimates were in the wrong direction, requesting additional data would indeed have been a likely outcome. Despite divergent outcomes in the primary endpoint, there were some key

similarities between the two studies. The low-dose aducanumab treatment arms, while not statistically significant, demonstrated consistent numerical effects favoring aducanumab in the studies. Also, aducanumab produced a time and dose-dependent reduction in brain amyloid burden. Upon closer review of the individual studies, Study 302 appeared to be strongly positive and internally consistent. In the context of a positive Study 302, the suggestion of a dose-response relationship observed in an earlier proof of concept study (Study 103), and the numerically favorable results of similar magnitude in the low-dose groups in both studies, the high-dose arm in Study 301 tends to stand apart not only for the negative outcome on the primary endpoint, but also the difference in biomarker profiles compared to Study 302. The estimate of the treatment effect in the high dose arm in Study 301 also showed the greatest change from the futility analysis to the final analysis, suggesting instability in the estimate.

Slide 18

Given the possibility that aducanumab is an effective drug for a disease with an enormous unmet medical need, FDA recognized that additional work was necessary to achieve a maximum understanding of the results and therefore established a collaborative plan with the applicant for further rigorous analyses of the data. By their nature, these analyses are post hoc and exploratory and therefore carry with them the

appropriate caveats and caution in their interpretation. To address these concerns, any exploration of the data was to be rigorous, limited in scope, and based on well-defined hypotheses. To the maximum degree possible, the analyses were pre-specified. An important distinction is that these analyses were not aimed at obtaining independent support from Study 301. Study 301 remains a negative study. The purpose of these analyses was to provide maximum understanding of the partially discordant results in the two studies and to determine if this understanding precludes independent consideration of Study 302.

Slide 19

Four key areas for further investigation were identified: differences in demographics and baseline disease characteristics, influence of ARIA and ARIA management, influence of study participants with rapid disease progression and the influence of dosing. A first step to establish the foundation for future investigation was to rule out differences in the characteristics of patients enrolled in the two studies, especially in the high dose arms. Differences were not expected because the protocols were identical, and the studies were run concurrently. And in fact, no differences in demographics were observed between the studies.

Slide 20

The motivation for investigating the influence of ARIA and ARIA management is its potential to cause functional unblinding of investigators, patients and caregivers because ARIA is associated with aducanumab treatment, is dose-dependent, and prompts differential management of patients. ARIA management also evolved over the course of the studies. For these reasons, functional unblinding was

considered a potential explanation for the positive results observed in Study 302. An indication of a systematic bias due to functional unblinding would undermine the strength of Study 302.

It is important to note that steps were taken in the protocol to minimize functional blinding, specifically the use of an independent rater who was blinded to patient management, including the occurrence of ARIA and whether the dose was modified. The FDA reviewed the protocols in 2015 and issued SPA agreements, indicating concurrence with their adequacy.

Slide 21

The incidence and radiographic severity of ARIA was generally similar between Studies 301 and 302, including the high dose arms, and is therefore not likely to inform the discordant results between the two studies. To address the potential effect of functional unblinding due to ARIA, the results of the primary analysis using all data were compared to the results of the same analysis using a reduced dataset in which assessments after the occurrence of ARIA were excluded. The results for CDR-SB are presented in this slide. For the low-dose and high-dose arms, approximately 30% and 40% of the observations were excluded, respectively. Overall, the results do not suggest a systematic bias due to functional unblinding. Similar results were also found for the other endpoints MMSE, ADAS-Cog 13 and ADCS-ADL-MCI. These analyses were also performed in subsets defined by dose, study and ApoE ε4 status and again no indication of a systematic bias was observed.

Slide 22

Investigation of individual treatment responses, particularly identification of patients with a rapid rate of disease progression was an area of focus because diagnostic plots of the primary endpoint analysis demonstrated that the distribution of the change from baseline in CDR-SB was skewed. This right-skewness is illustrated in the histogram on this slide. At the time of the futility analysis, the high-dose treatment arm was unique in that aducanumab-treated patients had greater cognitive decline than patient receiving placebo. A credible indication that aducanumab treatment was accelerating decline in Study 301 would undermine the results of Study 302. It was therefore also particularly important to interrogate the nature of the apparent decline in high-dose treated patients in Study 301.

Slide 23

A cutoff of an 8-unit increase in CDR-SB over 78 weeks was initially chosen to define patients with rapid progression. This increase is 4 times the expected estimate of mean progression used to power Studies 301 and 302. The numbers of patients in each study who met the cutoff were small, and evenly distributed across the treatment groups except for the high-dose group in Study 301 as seen in the table on this slide. Analyses for the primary and secondary endpoints were conducted using a dataset with all rapid progressors removed and compared to the results from the ITT population. Results for these analyses were only sensitive to rapid progressors in the high-dose treatment arm in Study 301. For example, the treatment effect estimate for CDR-SB went from 0.03 (or 2% in favor of placebo) to -0.09

(or 6% in favor of aducanumab). The analysis was repeated for other cutoff values to define rapid progression and the results were similar.

Slide 24

Potential group- and individual factors that may have contributed to the rapid decline in the 31 patients defined as rapid progressor were investigated but a distinct pattern was not observed. Even though the high-dose arm in Study 301 was the only aducanumab-treatment arm with a relatively higher number of rapid progressors, it was important to investigate whether aducanumab treatment itself was the cause of rapid progression. Rapid progression was evident at early time point in the study and investigation of ARIA and dosing patterns in patients with rapid progression did suggest a link with aducanumab treatment.

Taken together, this exploratory analysis suggests that small imbalances in the number of rapid progressors can have a relatively large impact on treatment effect estimates for the clinical endpoints, including when looking at subsets of data, and the high-dose arm in Study 301 was disproportionately affected by such an imbalance. With rapid progressors removed the high-dose aducanumab arm in Study 301 suggests a numerical trend in favor of aducanumab, consistent with every other aducanumab treatment arm. It is also important to note that 8 of the 9 rapid progressors from the high-dose arm in Study 301 were included in the futility data cut, further casting doubt on the appropriateness of the futility determination.

Now, let me be clear about what this analysis is not suggesting. It is not suggesting that rapid progressors should be ignored or routinely removed from analyses of clinical trials. The existence of rapid progressors has been noted in the literature and similar patients have been observed in other clinical studies. Also, this analysis is not suggesting that Study 301 was a negative study because of a few outliers. This analysis is simply intended to shed light on the discordant results for the high dose in the two studies. A small imbalance in rapid progressors simply explains a considerable portion of this difference.

Slide 25

Dosing was an obvious area of investigation to understand the discordant results for the high dose arms. The lack of adequate dosing has been cited as a contributing factor to the failure of previous clinical trials of amyloid-targeting therapies. Also, protocols of notable clinical trials have been amended in recent years to significantly increase dose levels. The importance of dose was also directly established for aducanumab in the proof of concept study 103. The motivation to investigate the influence of dosing in the interpretation of Studies 301 and 302 is further supported by the conduct of these studies and, in particular, amendment that were made to the protocol. Specifically, protocol version 3 modified ARIA management rules so that patients were more likely to be able to continue dosing or achieve target dose levels. Protocol version 4 increased the target dose for ApoE ε4 carriers in the high dose arm from 6 mg/kg to 10 mg/kg.

One challenge in comparing dosing over time or between studies is identifying an appropriate dosing metric. Steady-state is achieved after 4 consecutive doses and dose interruptions can have long-term consequences on amyloid reduction. There was also heterogeneity in the dosing profiles due to changes in the protocol and dose modifications due to adverse events. It may be possible that achieving consistent exposures early in the study is more important than reaching the target dose later in the study. It is also worth noting that patients receiving lower doses had more consistent dosing compared to patients receiving higher doses. For these reasons, multiple measures of dosing were characterized, including cumulative dose, number of doses at target dose, and number of uninterrupted doses at target dose. Dosing is also confounded by other factors, including ARIA and ApoE ε4 carrier status.

Slide 26

The bar plot on this slide illustrates the dosing patterns over time for Studies 301 and 302. The height of each bar in this plot represents the median number of 10 mg/kg doses received for each subset of enrolled subjects. This plot demonstrates two key points. First, the timing of the protocol amendments was such that patients enrolled later in the studies were more likely to achieve higher and more consistent aducanumab exposures. Second, the timing of the studies and pace of enrollment was such that patients in Study 302 were more likely to benefit from changes to the protocol than patients in Study 301. The difference in dosing between Study 301 and Study 302 was less than the difference in dosing between early enrolled patients and later enrolled patients.

Slide 27

A guiding principle of this investigation was that if aducanumab is effective and the effect is dose-related as in Study 302, it follows that patients in Study 301 with adequate and consistent dosing should also demonstrate an effect on clinical endpoints. An absence of an effect in this subgroup of patients in Study 301 would diminish the persuasiveness of Study 302. Although it is impossible to fully account for all factors that may contribute to findings in subgroups formed by post-randomization factors such as dosing, a variety of approaches, each with strengths and limitations, were employed. This slide shows the results from one of the approaches. Each panel in this plot shows the treatment effect for three different exposure subgroups within the high-dose treatment arms of Studies 301 and 302. Subgroups were chosen to have an approximately even number of subjects within the three groups. Each aducanumab-treated patient was matched with a placebo patient based on key baseline demographic and disease characteristics, including ApoE ε4 status, region, sex and time enrolled in the study. First I would like you to focus on the panel on the right which shows patients who had at least 8 consecutive doses of 10 mg/kg at steady state. This panel shows that patients in Study 301 who had consistent exposure to the target 10 mg/kg dose had a favorable treatment response of similar magnitude as patients in Study 302. It is worth noting that in Study 302, a higher percentage of patients had this level of exposure compared to Study 301, 35% to 29%. Now I will talk about the panel on the left which shows results for patients in the high-dose arm who did not receive any of the 10 mg/kg target doses. In both studies the effect appears to be not much different from placebo. This time Study 301 had a higher percentage of patients with 0 doses of 10 mg/kg, compared to Study 302, 29% vs. 24%. The middle panel shows patients with heterogeneous dosing, such that they received anywhere from 1-7

consecutive doses of 10 mg/kg at steady state. A discrepant result is clearly seen here with the treatment effect favoring placebo in Study 301.

Slide 28

The figures on this slide show the results of the same analyses, but with the rapid progressors removed. Here we can clearly see that rapid progressors account for some of the discrepancy observed in the middle panel for the middle exposure group, but not all of it. It is worth noting that this middle group includes the most heterogeneity in terms of dosing profiles. Even with this heterogeneity, pharmacokinetic and PET data indicate that the aducanumab exposure and amyloid reduction in this subgroup is in excess of the low-dose arm in the study. Therefore, if one accepts that the low doses in Studies 301 and 302 consistently demonstrate a numerical effect in favor of aducanumab, one would expect to observe at least a numerically favorable outcome in this intermediate exposure subgroup similar to the effect observed for the low dose arms. Given the multitude of findings suggesting a dose-response relationship and a directionally favorable effect of aducanumab, the fact that this is not the case in this one subset raises the possibility that this observation may simply be due to chance or some other unknown factor unrelated to aducanumab.

Slide 29

Overall, the analyses on the influence of dose indicate that dosing in an important consideration for interpretation of the efficacy results in the two studies. In general, consistent exposure to the target dose of 10 mg/kg was associated with better clinical outcome in both studies and patients in Study 301 who had higher exposure to the 10 mg/kg dose demonstrated treatment effects similar in magnitude to patients in Study 302. Although the precise relationship between dosing and treatment effect is unknown, the difference in various measures of aducanumab exposure between the studies is modest and dosing alone does not explain the negative finding for the high dose in Study 301. Also, there remains a subset of patients in Study 301 who received intermediate exposure to 10 mg/kg yet failed to show a treatment effect of similar character to patients who received intermediate exposure to 10 mg/kg in Study 302 or even subjects who received the low dose in either study. For these reasons, these analyses do not provide independent evidence of the effectiveness of aducanumab, but rather contribute to the overall understanding of Study 301.

Slide 30

It may be tempting to misinterpret the considerations presented in this part of the presentation as “explaining why Study 301 was negative” or providing a full account of the differences between the studies. This was not necessarily the intention of this work and such an explanation is not necessary to establish the effectiveness of aducanumab. At the June 14, 2019, Type C Meeting, the Division clearly stated that, “available data do not suggest the future use of Study 301 as an efficacy study providing independent evidence of effectiveness supporting the approval of aducanumab.” Rather, the Division noted that analyses “may have a role in supporting the results of Study 302,” or “may be understood

well enough… to not represent evidence that the drug is ineffective." The analyses presented here are exploratory by design but limited in scope and focused on pre-defined areas of interest. The rapid progressor analysis indicated that a small imbalance in the number of rapid progressing patients in the high-dose arm in Study 301 had a disproportionate impact on the estimate of the treatment effect using the primary analysis method. An examination of dosing in Study 301 indicates that patients with higher exposure to the 10 mg/kg dose in Study 301 had similar responses to patients in Study 302. These two factors contribute to the overall understanding of Study 301 and together indicate that the results of Study 301 do not meaningfully detract from the persuasiveness of Study 302. There were no findings from the exploration that represented evidence that aducanumab is clearly not effective.

Slide 31

The partially discordant results in Studies 301 and 302 elevate the importance of considering the results of Study 103. Although designed primarily as a safety and tolerability study, Study 103 was an adequate and well-controlled study that explicitly included assessment of prespecified clinical and biomarker endpoints and a prospectively identified statistical analysis plan. The study informed the design of Studies 301 and 302 but also included similar elements as these studies, including the requirement of a positive amyloid PET scan and blinded assessment of clinical endpoints. Notable differences include the smaller sample size, inclusion of patients further along the disease continuum, and enrollment of patients only in the United States.

Slide 32

Statistically significant reductions in brain amyloid were achieved for all treatment groups at Week 54 with a clear dose- and time-dependent relationship as illustrated in the figure on this slide. The magnitude of the reduction in brain amyloid plaque observed with the fixed-dose of aducanumab 10 mg/kg at Week 54 in Study 103 is almost identical to that seen with the high dose of aducanumab in Study 302 at Week 78 as shown in the table. The number of 10 mg/kg doses to be administered in the placebo-controlled period, 14, was also the same in the two studies. Therefore, the 10 mg/kg fixed-dose arm in Study 103 is the relevant arm to compare to the high dose arm in Study 302.

Slide 33

Even though the study was not powered to detect changes in clinical endpoints, a statistically significant treatment effect was demonstrated in the 10 mg/kg fixed-dose group as compared to placebo on both the CDR-SB and the MMSE as indicated by the shaded cells in the table. In general, a dose-dependent relationship for clinical endpoints was also observed. The clinical efficacy results in Study 103 are supported by the statistically significant, dose-dependent reduction in brain amyloid plaque as illustrated on the previous slide.

Slide 34

It is important to note that there are design and analysis limitations in Study 103 as compared to a typical confirmatory Phase 3 study design. In particular, because of the staggered dose design and randomization scheme there is no direct concurrent randomization to the various treatment arms to inform a dose-response analysis or a dose vs. placebo comparison based upon concurrent randomization. However, the placebo arms were pooled across cohorts in compliance with the prespecified statistical analysis plan because of the uneven randomization to placebo in each cohort within the staggered design. Although not prospectively controlled for multiplicity, the choice of endpoints and analytical approach is consistent with that which would have been anticipated should an analytical hierarchy have been in place, and that prespecified elements were respected.

Slide 35

Based on the considerations described in this presentation, I have concluded that the applicant has provided substantial evidence of effectiveness to support approval. In summary, Study 302 provides the primary evidence of effectiveness as a robust and exceptionally persuasive study demonstrating a treatment effect on a clinically meaningful endpoint and reinforced by effects on secondary endpoints, biomarkers, and in relevant subgroups. Study 103 was an adequate and well-controlled study which included design components consistent with Study 302 and, for the reasons I have described earlier, I have concluded demonstrated a persuasive treatment effect on both clinical endpoints. The dose-response relationship for amyloid beta reduction provides support for the positive finding in the 10 mg/kg treatment arm and to the apparently dose-related effects observed on clinical outcomes in Studies 103 and 302. Study 301 does not contribute to the evidence of effectiveness. The results of exploratory analyses, however, contribute to the overall understanding of Study 301 and together do not meaningfully detract from the persuasiveness of Study 302.

November 6, 2020 BLA 761178: Peripheral and Central Nervous System Drugs Advisory Committee Meeting

Script for FDA Presentation: Aducanumab for the Treatment of Alzheimer's Disease: Clinical Overview of Safety

FDA Presenters: Natalie Branagan, MD and Brian Trummer, MD, PhD

Slide 2

I am Dr. Natalie Branagan, Medical Officer in the safety team of the Division of Neurology 1. In this presentation, Dr. Trummer, Medical Officer in the Division of Neurology 1 and I will discuss the safety of aducanumab, with an emphasis on Study 302. The safety findings from Study 302 are consistent with those presented in the briefing package for Pool A1, which combines Studies 301 and 302. I will discuss the overall safety findings of Study 302, including deaths, serious adverse events, and most common adverse events from Study 302. Dr. Trummer will then discuss the safety findings related to ARIA, and I will present our conclusions to date regarding risk mitigation.

Slide 3

This slide shows the incidence of death among all subjects in Study 302. The incidence of death in aducanumab-treated subjects in Study 302 was not in excess of placebo.

Overall, across the clinical program, there were 31 deaths. There was not an excess of deaths in aducanumab-treated subjects compared to placebo in Study 103 or in pooled Studies 301 and 302. Most subjects had underlying risk factors for events with fatal outcome. No deaths were attributable to treatment with aducanumab.

Slide 4

The following slides will focus on the 10 mg/kg dose.

The overall incidence of serious adverse events was similar in the aducanumab 10 mg/kg group and in the placebo group.

The most frequent serious adverse events included ARIA-E, hemorrhage, malignancy, inguinal hernia, and superficial siderosis of the central nervous system.

Slide 5

This slide shows the most frequently reported adverse events occurring with an incidence of at least 2% higher in the aducanumab 10 mg/kg arm compared to placebo. The most frequently reported adverse event was ARIA-E, which was reported in 34% of subjects in the aducanumab 10 mg/kg arm compared to 2% of subjects in placebo. ARIA-H was reported at an incidence of 20% in the aducanumab 10 mg/kg arm compared to 7% in placebo. Superficial siderosis of the

central nervous system was reported in 13% of the aducanumab 10 mg/kg arm compared to 3% of placebo. Other common adverse events included headache, dizziness/vertigo, confusion/delirium/altered mental status/disorientation, diarrhea, nausea/vomiting, visual disturbances, and nasal congestion.

Dr. Trummer will now present safety findings related to ARIA.

Slide 7

Hi, I am Dr. Brian Trummer, Medical Officer in the Division of Neurology 1. ARIA, which stands for Amyloid Related Imaging Abnormalities, is a phenomenon known to occur with monoclonal antibody therapy against amyloid-beta. There are two types of ARIA: vasogenic edema (ARIA-E), and hemorrhagic ARIA-H, which consists of cerebral microhemorrhage, superficial siderosis, and cerebral hemorrhage. As shown in this table, ARIA was reported as a serious adverse reaction in 1.5% or 8/547 subjects in the aducanumab 10 mg/kg group, compared with 0.2% in patients on placebo. All 8 of these subjects experienced ARIA-E. Three of the subjects also had ARIA-H superficial siderosis, 2 also had ARIA-H microhemorrhage, and 1 also had an ARIA-H cerebral hemorrhage. Symptoms experienced in these subjects with serious adverse events included confusion, disorientation, delusional thoughts, seizure, agitation, gait disturbance, right hand weakness that resolved in 12 days, decreased coordination, ataxia, a vision disturbance characterized as seeing little black and white spots and smooth surfaces appearing rough that resolved in 66 days, headache, nausea, a visual disturbance that resolved in 80 days, falls, and blurred vision that resolved in 14 days. One subject died, but the case was not related to ARIA.

Slide 8

This table illustrates the frequency of ARIA in the placebo controlled period of Study 302. Overall, ARIA occurred in 41.7% of those in the 10 mg/kg aducanumab group compared to 10.2% of those in the placebo group. All subtypes of ARIA were increased in the treated group relative to the placebo group, although cerebral hemorrhage was rare and differed minimally from placebo. The majority of ARIA was asymptomatic, and a purely radiological finding. Symptomatic ARIA occurred in 7.5% of the 10 mg/kg aducanumab treated group. The most frequent symptoms attributed to ARIA and experienced by more than 1 subject were headache, confusional state, gait disturbance, blurred vision, dizziness, fall, fatigue, visual impairment, and balance disorder.

Slide 9

In conclusion, the safety profile of aducanumab for the treatment of Alzheimer's disease has been adequately characterized and is acceptable for the proposed indication.

ARIA can be managed by appropriate labeling language, which would include a warning in the prescribing information.

Prescriber and patient education regarding ARIA would be important, and appropriate methods to communicate and educate healthcare providers and patients about the risk of ARIA are under consideration

Title BLA 761178 Statistical Review Aducanumab in Alzheimer's

Slide 1 Title Slide

Slide 2 Acronyms used in the Slides

Slide 3

The phase 3 study planned lower doses for the APOE carrier stratum due to expected increased risk of amyloid imaging related abnormalities as seen on MRI imaging and previously experienced with other related Alzheimer's candidates. The phase 3 studies were started before the phase 1b safety study, study 103 was completed. The studies were 78 week parallel group trials with 545 patients per group and 1:1:1 randomization to placebo low or high dose stratified by APOE carrier status and Site. Approximately 2/3 of the study population are APOE carriers. When study 103 results were available in early 2017 the sponsor felt based on that data that it would be safe to increase the dose in apoe carriers to agree with the high dose in non-carriers of 10 mg/kg. Thus protocol amendment 4 was written in march 2017 and allowed for only the high dose carriers to increase the dose to the maximum already utilized in non-carriers. Non carriers had a consistent maximum for low and high groups of 6 and 10 mg/kg respectively . In carriers the low dose had a consistent maximum dose of 3 mg/kg but the high, which started with a maximum of 6 in carriers, was increased to 10 mg/kg mid-study with the implementation of protocol amendment 4.

Slide 4

The phase 3 study planned lower doses for the APOE carrier stratum due to expected increased risk of amyloid imaging related abnormalities as seen on MRI imaging and previously experienced with other related Alzheimer's candidates. The phase 3 studies were started before the phase 1b safety study, study 103 was completed. The studies were 78 week parallel group trials with 545 patients per group and 1:1:1 randomization to placebo low or high dose stratified by APOE carrier status and Site. Approximately 2/3 of the study population are APOE carriers. When study 103 results were available in early 2017 the sponsor felt based on that data that it would be safe to increase the dose in apoe carriers to agree with the high dose in non-carriers of 10 mg/kg. Thus protocol amendment 4 was written in march 2017 and allowed for only the high dose carriers to increase the dose to the maximum already utilized in non-carriers. Non carriers had a consistent maximum for low and high groups of 6 and 10 mg/kg respectively . In carriers the low dose had a consistent maximum dose of 3 mg/kg but the high, which started with a maximum of 6 in carriers, was increased to 10 mg/kg mid-study with the implementation of protocol amendment 4.

Slide 5

On this slide we see the primary endpoint results for the two phase 3 studies as well as a pooled study analysis on the right reported by the sponsor. The high dose primary p-values are .83 for study 301 and .01 for study 302. The prespecified multiplicity adjustment for multiple doses and multiple key secondary endpoints called for a sequential closed testing for each endpoint. This prioritizes testing the low dose on the primary endpoint before testing secondary endpoints for the high dose. Prioritization is needed to control the experimentwise type I error over all prespecified key hypotheses at .05 2 sided. Because the low dose did not win on the primary endpoint in either study, with this multiplicity plan the possibility of claiming formal significance of the high dose on the key secondary endpoints is lost.

However, the secondary results are displayed to allow consideration of the full data package. The highest priority secondary endpoint p-value was originally .06 in study 302, not technically significant, but when slightly more data was later collected it's pvalue decreased to .049. The lower prioritized secondaries have nominally significant p-values in study 302 but would only be interpretable if all higher prioritized endpoints were significant first, for both doses. The reviewer found that the Secondaries endpoints are moderately highly correlated with the primary (pairwise correlations ranging from .40 to .64) which suggests that they are not measuring very distinct clinical contructs/information and we can observe that within both studies the four endpoints were all consistent in terms of the nominal significance of their p-values ,i.e., all negative in 301 and all positive in 302 (for MMSE after the late data addition). IT seems important to note that the high and low dose share the same placebo group as comparator so these differences are correlated. If the placebo worsens then both high and low dose would tend to improve in comparison all else being equal. This also applies to overall trends for the low dose across the studies, neither of which was statistically significant.

Slide 6

Both Phase 3 Trials (301,302) for Aducanumab were futility stopped at 50% completion after meeting the futility stopping criteria that utilized study-pooled treatment group estimates. Based on study specific treatment group estimates at the interim the high dose conditional power at the interim was 58.6%. It seems unlikely that a futility stopped trial could produce a clinically meaningful effect. In that analysis the estimated treatment difference from placebo at Week 78 was -.28. In the final ITT analysis the high vs. placebo difference at Week 78 on the primary endpoint increased to -.50. Interestingly, the low dose was numerically better than the high dose in the post-interim ITT patients. This seems to challenge the importance of 10 mg/kg since 80% of this high dose group had access to the dose increase from 6 mg/kg to 10 mg/kg dose in carriers under protocol amendment 4 and yet the low dose was numerically better and these two groups are balanced by randomization . The low vs. high difference is even more striking in those who had the opportunity to complete to week 78 before the futility stopping. A bigger effect observed in data after an interim is a concern for operational bias potentially creeping into the study conduct from the interim unblinding of data. There were firewalls planned but the existence of such a trend means it is more difficult to rule out operational bias due to interim unblinding. More than 40% of the ITT population did not have an opportunity to complete the study to Week 78, the primary timepoint, due to the futility stopping. The result in those who did have the opportunity to complete had a p-value of .0368 in study 302. The apoe carriers only got access to the purported effective higher high dose of 10 mg/kg after protocol amendment 4 . The average follow-up of those with access to 10 mg/kg to the maximum extent allowed in protocol 4, 14 doses, is about 50 weeks rather than 78, so there is also less safety follow-up on 10 mg/kg in APOE carriers.

Slide 7

There were 2 waves planned for a collaborative workstream that was started between the Division and the sponsor in July 2019 to seek to understand the conflicting phase 3 results, given the unmet need in early Alzheimer's. In the first wave , a plan was drawn up to complete many simulated complete trials using the method of multiple imputation, in which a model for missing outcomes is used to impute missing outcomes for those missing the Week 78 outcome by applying the model to data from completers with similar baseline demographic and baseline disease characteristics and generating predictions for the missing outcomes. Many simulated completed trials were generated and

summarized. The result of this work confirmed that study 302 would likely have been statistically significant if it had not been futility stopped but it should be noted that at the time the FDA had only recently received the data and had not performed a full BLA review and had the data for less time than the sponsor. Simultaneously, however the wave 1 work suggested that study 301 would likely still have failed. Furthermore, simulated complete trials are always more uncertain than a truly completed trial.

The second wave of collaborative work sought to understand any important differences between study 301 and 302 that may have led to the difference in study results. A clinical question of interest was whether there might be 302 like patients in study 301. This second wave became more centered around pharmacokinetic and pharmacodynamic data analyses and therefore the statistical team became less involved.

Slide 8

Here we see the Study 302 CDRSB change from baseline profile over time for each treatment group at the 3 post-baseline visits at which CDRSB was assessed. The numbers in the legend reflect the ordered dose group, 0 or blue is placebo, 1 or red is low dose and 2 or green represents the high dose group. We can see that the vertical lines are not separated until Week 78, i.e., there were no significant treatment effects before Week 78. It is also notable that because of the futility stopping more than 45% were missing the Week 78 CDRSB in the final ITT analysis, as indicated by the sample sizes denoted next to the visits along the horizontal axis.

Slide 9

Study 301 fell dramatically short of study 302, and provides a stark contrast and challenge to the 302 results. The high dose was numerically worse on average than placebo at Week 78 in study 301. After examining the data the sponsor developed a theory that so called rapid progressors in study 301 undermined the high dose because there were more extreme poor outcomes in study 301 than 302, **even** in the study 301 high dose . The idea of rapid progressors has become an often cited challenge to success in rare disease clinical trials. Alzheimer's is far from rare but there is a dominantly inherited type that is rare. A platform trial was recently conducted in dominantly inherited Alzheimers investigating multiple candidates to increase efficiency of use of resources in the setting of a rare form of Alzheimer's. The analysis for that trial did not discriminate between normal and rapid progressors despite having less ability to increase sample size in case of their existence. Again, the rapid progressors theory for the failure of study 301 was only presented as a challenge after data unblinding for Aducanumab. Study 301 and 302 actually had blinded sample size increases from 450 to 535 patients per group which should have helped if rapid progressors are real. The sponsor otherwise describes this treatment as disease slowing but suggests it failed in study 301 partly due to rapid progressors in early Alzheimer's, meanwhile in study 302 the high dose shows more effect in the Mild AD subgroup rather than in the complementary prodromal or presymptomatic baseline disease stage. Various analyses conducted by this statistical reviewer suggest that the rapid progressors in the high dose of 301 may be more of an indication of a systemic lack of efficacy.

For example, excluding several outliers only for the high dose does not make the high dose significant or even numerically better than the low dose, which it can be compared to on the basis of the 1 to 1 to 1

randomization. Two Statistical methods , robust regression and trimmed means, that are more designed for reliability in the presence of outliers than the prespecified primary analysis method also did not suggest a significant effect of the high dose in study 301. Since 301 is a large study this showed that the influence of any individual outlier patient is limited and the primary analysis of study 301 is legitimate and the conflicting phase 3 results issue remains.

Slide 10

The two figures on this slide show mean CDRSB profiles over Visits by treatment group and by pre vs. post PV4, that is before and after the mid-study high dose dose increase, within treatment group for study 302 on the left and 301 on the right. This is all for the amendment 4 affected carrier stratum i.e., APOE+ stratum only, since it was the only stratum affected by the mid study high dose increase. This stratum accounts for just over 2/3 of the overall study populations. Study 302 exhibits a dramatic worsening in placebo post-PV4 , protocol amendment 4, compared to pre-PV4 in the stratum with the high dose increase, but no improvement of high dose post-PV4 compared to pre-PV4 high dose is apparent or high dose to low dose post pv4 despite PV4 implementation of the dose increase for the high dose. We see in the leftmost subfigure for placebo that the pre-PV4 in blue and post-PV4 in red placebo subgroup's profiles are totally separated across all visits. In the 301 half of the slide on the right, while the high dose on the left improved post pv4 relative to pre-pv4 it is still comparable to the low dose post-PV4 group and the placebo post-PV4 also worsened pre to post-PV4. The overall numbers, i.e., including the non-carrier stratum, are also shown below the figures and these numbers also show a worsening of placebo post-PV4 compared to pre-PV4 (see the yellow highlighted columns).

The figures imply that the Effect of the Dose Increase is entangled with Simultaneous Placebo Worsening and it is not possible to disentangle cause and effect.

Slide 11

The sponsor argues that intermediate dosing that is less 10 mg/kg doses in APOE carriers may have stopped 301 from being significant since it was slightly ahead of study 302 in the timeline and thereby may have gotten less benefit from the late dose increase. However, this glosses over the fact that the non-carriers who comprise a 1/3 of study population got 10 mg/kg from the start of the study and the following slides will show they have less effect in study 302. This would suggest either a treatment by APOE interaction, that is systematic less effect of the drug on non-carriers or it challenges the sponsor's intermediate 10 mg/kg dosing theory, because those with the consistent most 10 mg/kg doses showed very little effect on average in study 302. The high dose non carriers substratum is numerically worse than placebo in study 302 on the first key secondary endpoint, MMSE and the test for interaction for this comparison has a p-value of 0.07. An interaction test of APOE by Treatment for all 3 groups has a pvalue of 0.0065 for MMSE. The opposite group of APOE positive, carries study 302 but is in the wrong direction in 301, numerically worse than placebo, so again there are no unconflicted results or replication of positive effects across the two studies. It seems worrisome that the group driving the apparent efficacy in 302 had a mid study dose increase and more unblinding and trended in the wrong direction in 301. and the whole pattern of study outcomes could be explained by a randomly worsened placebo after the dose increase amendment in study 302.

Slide 12

The figure shows that the high dose non-carrier group with highest dose throughout study, higher on average than carriers as shown on the right side, (and less ARIA adverse events) was numerically worse than placebo

on first key secondary and numerically less effect than carriers on all 4 key endpoints The blue is the APOE non carrier LS mean difference from placebo and the red is the APOE carrier LS Mean difference from placebo and moving down the left column the figures correspond to the prespecified order of the primary and key secondary endpoints. The dotted black line shows the zero difference level for reference. One can see that the second row MMSE center of the box is below zero, that is numerically worse than placebo. We can notice that for each one of the four key endpoints although the dose is higher on average as seen in the right column, i.e. more 10 mg/kg doses for blue, the blue effects for non-carriers are all smaller than those for the carriers in red. For the first key secondary a test for interaction between APOE and treatment is highly significant p=0.0065 and the trend across all four endpoints support this as less likely being a coincidence. Tests for interaction specific to the high dose vs placebo comparison have p-values of .07 for both MMSE and ADCSADL and .15 for CDRSB. They all suggest the same direction of interaction, i.e., non carriers have less high dose effect. But again the paradox if 10 mg/kg is effective is that non-carriers got more consistent 10 mg/kg doses from the start of the study than the carriers who only received 10 mg/kg after the mid-study protocol amendment but they show consistently less treatment effect.

Slide 13

This slide shows the same figure as the last slide but applied to the low dose vs. placebo comparisons in study 302 and it shows the exact same trend as for the high dose which seems to add strength to the likelihood that the interaction is real (if there is a high dose effect) . In particular, for the low dose the non-carriers all trend in the wrong direction compared to placebo across the primary and key secondary endpoints although by protocol they had a higher dose, i.e., the low dose was 6 mg/kg for non-carriers vs. 3 mg/kg for carriers.

Slide 14

Here we summarize the extent of partial unblinding of individual patients to investigators and some limited sponsor personnel due to the treatment related adverse event of ARIA which often necessitated dose titration adjustments for patient safety. Very Limited sponsor personnel were unblinded but this shows that they were partially unblinded for a lot of patients since so many high dose APOE carriers required dose management due to ARIA,i.e., 35% . The figure shows the percentages with titration modifications for each treatment group broken up by carrier non-carrier and pre-PV4 vs. post PV4. It is clear that it was much higher in Aducanumab groups and in carriers compared to non-carriers. It is important to consider whether this could have affected the study conduct or the analysis results. The primary efficacy measure in this study is a physician rating that could potentially be susceptible to bias if the treatment was unblinded to the physician , patient or caregiver and there was increased MRI monitoring for affected patients after ARIA events. It turns out that Although excluding primary endpoint data for patients unblinded due to ARIA is somewhat reassuring not suggesting overt sensitivity to this issue it Can't completely rule out bias due to ARIA unblinding by excluding cases. This is because exclusion or censoring of affected data for ARIA cases breaks the randomization and/or imbalances the treatment groups distributions of follow up (it's a post-randomization event defined group), especially for such a strongly treatment related adverse event.

Slide 15

The exploratory safety study 103 had a staggered design. The words in the boxes are not important, the point of the figure is to illustrate the staggered parallel groups design . The important thing to notice is that the cohorts were spaced out in time and that the sponsor's analysis combines 3 placebo groups that had no chance of randomization to 10 mg/kg ,Arm 4, indicated by the red arrow with one that did (Arm 5). The overall randomization does not directly support this since the cohorts had separate randomizations. Arm 9 was all APOE positive which increases the likelihood of an APOE imbalance between 10 mg/kg and the pooled placebo.

Slide 16

Study 103 was also shorter than 302 , with 54 weeks rather than 78 week follow up. Likely due to the staggered arms and pooling of placebo arms there were moderately more females and more APOE carriers in pooled placebo than 10 mg/kg group. Although the week 54 difference of 10 mg/kg vs. pooled placebo has a p-value of .0436, this comparison loses significance if data affected by the intercurrent event of post-baseline starting of concomitant Alzheimer's medications is censored. Both pooled placebo and 10 mg/kg groups had cases of post-baseline starting of concomitant symptomatic Alzheimer's medications. The randomization backed analysis of 10 mg/kg against the concurrent placebo , arm 5 only, is not significant p=.12. Given the lower effects observed in phase 3 for the APOE non-carriers APOE subgroup effects were examined for this study and the opposite was found , the effect in non-carriers was almost twice that of carriers, a stark contrast from phase 3. Thus this study does not replicate the effect in APOE carriers of study 302 and 302 does not replicate this study 103 suggestion of effect in non-carriers, which is 20 times larger than the estimate in 302. Also as in the phase 3 study , 302, There was also no suggestion of an effect before the last visit in study 103.

In summary this study was an exploratory safety study. the efficacy results are Not robust. This is a much smaller study without randomization directly backing the sponsor's analysis so there is no reason this data could override a large well controlled failed trial , 301, especially outside of the rare disease setting.

Slide 17

Let's return to the phase 3 trial data. The sponsor argues that there are study 302 like patients in study 301. In particular the subset with the full 10 mg/kg dosing. However, this is a post-randomization event defined subgroup so it has no proper control. That is, it is not clear which subset of the post-PV4 placebo group would have been most compliant and free from ARIA **if** they had been assigned to the high dose. We can't just compare the full 10 mg/kg dosed subset to all placebo because one group will have selection bias compared to the other. The sponsor tried to match placebo patients to the actual achieved dosing subset of the high dose by a post-hoc propensity score model. However it is clear that the matching is inadequate since the arrows indicate that the placebo response trends worse as the number of 10 mg/kg doses increases. Therefore again this analysis is confounded with a non-constant worsening placebo effect as the dose increases. Rather than a better effect of the high dose in those with more 10 mg/kg doses it could just be that the placebo comparator subgroup was worse. This subset matched comparison could never equal or recreate a randomization. Changing demographics in later enrolled also confound the interpretation of the effect of raising the dose and actual doses (different regions, more MCI, etc.). Even for those with the opportunity to receive 14 doses of 10 mg/kg, only 30-40% actually did, so tolerability seems to be an issue . The proper MRI monitoring for ARIA may be more

challenging to achieve in a non-controlled setting outside of a clinical trial, particularly in the current pandemic. In APOE non-carriers, who are less susceptible to ARIA than carriers, 61% in study 302 were able to have 14 doses of 10 mg/kg yet they showed very little effect , -.06 points difference from placebo on Week 78 CDRSB.

Slide 18

One possibly underappreciated difference between 30 1 and 302 is in the enrolled countries. There were 6 countries unique to 302 and 6 others unique to 301 and 7% more US contribution to 301 than 302.

There was also a significant Country by Visit interaction and Country by Visit by Treatment interactions in study 302, indicating that CDRSB change profiles over Visits were significantly varied by Country and treatment differences also varied significantly by country. This raises the possibility that country differences could explain at least part of the difference in outcomes between 301 and 302. There was also a change in distribution of country enrollment after protocol amendment 4, with for example a decrease in US enrollment and increasing Asian enrollment. In study 302 enrollment in a country favoring placebo numerically, Poland, decreased after PV4 from 19 to 6%, while enrollment in Japan which favored the high dose increased post PV4 from 2 to 12%. Given the inconsistent CDRSB profiles across countries and change in enrollment by country the effect of the dose increase can't be isolated. In addition to change in distribution of country enrollment after PV4 there were also moderate changes in enrollment of baseline disease stages and use of symptomatic AD medications at baseline from before the amendment to after. These changes further confound the effect of the protocol amendment mid-study dose increase since these variables, like country, also significantly affected the CDRSB changes profile over visits when added to the analysis model.

Slide 19

Here we have a forest plot to examine high dose effect at Week 78 compared to placebo on the primary endpoint for various subgroups of importance. If one was to rely on only one positive study one would hope to see a high degree of consistency within the study across subgroups. However, it appears that the high dose effect is inconsistent across age subgroups, with higher effect in the oldest subgroup and virtually no difference below age 70. The estimated effect is much larger in the Mild AD baseline stage as compared to the Prodromal AD baseline stage, which was perhaps not unexpected given the inclusion of this effect as an explanatory variable in the primary analysis model. One might presume that a truly disease slowing treatment may if approved be used in practice in a preventive manner targeting younger presymptomatic patients with risk of AD but the evidence for Aducanumab in the younger and more pre-symptomatic patients is lacking compared to older and more symptomatic patients as shown here. As previously mentioned in detail, there was also a much larger effect in APOE carriers. The variability in treatment effects by Country is also displayed and at the bottom we see the 301 effect in the US was in the wrong direction so the high dose effect in the US was also inconsistent across the two phase 3 trials. Furthermore, the APOE carriers drive study 302 since they are the larger stratum and show a bigger effect than non-carriers in study 302, yet we see at the bottom that the 301 effect in carriers was numerically favoring placebo, so the 302 effect in carriers is challenged by the results for the same stratum in 301.

Slide 20

Aducanumab showed statistically significant dose dependent changes from baseline in amyloid beta plaques in the cerebellum, designated as the primary brain region, as measured by PET SUVR imaging. This is not surprising given that the treatment is an amyloid beta agonist. The important question is do individual changes in amyloid actually correlate with or predict long term cognitive and functional changes as measured by CDRSB. If we assess this in the figure on the left within the high dose group there appears to be no correlation at all in study 302 as shown by the cloud of red dots and a superimposed fitted red linear relationship. The slope of the red line is near zero but slightly negative which is the wrong sign needed for a slope to support a link between this biomarker and clinical outcome. The blue dots show the same changes of CDRSB and PET SUVR at Week 78 for the failed study 301. Here the slope is slightly positive but the linear relationship is very weak as indicated by the wide spread of blue points around the fitted blue line. On the right a correlation within the sponsors post-hoc highlighted full 10 mg/kg dosed subgroup is assessed and again there is no correlation even in this maximally 10 mg/kg dosed subset. Furthermore, again it is technically in the wrong direction but essentially zero. Therefore, it is not clear that there is any linkage between reduction in plaque and long term clinical change. This seems to call into question whether there is truly any evidence of disease slowing of Aducanumab. There could also be untoward rare adverse effects of the dose dependent amyloid plaque removal that may not have been observed yet due to futility stopping of the trials.

Slide 21

The sponsor also has highlighted an exploratory change in tau plaque biomarker as measured in the medial temporal brain region by PET imaging. They showed a correlation between reduction in tau as a function of average dose as shown in the left figure. This is an exploratory endpoint in an extremely small sample of about 30 patients. Furthermore, 83% of this CDRSB data was collected after the futility press release, this data was censored in the primary analysis and all but one placebo and high dose group patients were from study 301, only one was from 302. The correlation was actually stronger in the low dose than the high dose. It turns out that the baseline CDRSB are not very similar between the selected placebo and high dose patients in this voluntary non directly randomized substudy differing by just under a half point which is larger than the 302 high dose effect at Week 78. The drug does not even directly target tau and without a correlation between its target amyloid and clinical change at the patient level as we examined on the last slide, this correlation with an off target exploratory biomarker effect seems more dubious. Also, as seen on the right, there is no correlation within the high or low dose between reduction in tau and Week 78 clnical change as measured by CDRSB. In summary this is an exploratory analysis in a voluntary non-randomized sample and regardless of the trend in tau change as a function of dose there is no correlation within the high dose or low dose between changes in tau and change on CDRSB. So the tau reduction does not seem to be predictive of clinical change at the patient level. At best this relationship between tau reduction and average dose is an exploratory analysis and more robust evidence would be needed to support a disease slowing effect.

Slide 22

Benefit risk is an important consideration for a new molecular entity. The futility stopping of the phase 3 trials and mid-study increase to 10 mg/kg for APOE carriers implies that there is less long term safety data at 10 mg/kg in carriers who seem to drive the 302 efficacy than there would be if the trials had not been stopped for futility or the dose had not been increased in the middle of the study. The average follow up for those APOE carriers to which 10 mg/kg was fully available was only 50 weeks rather than

78. An increase in falling adverse events was observed in the high dose as compared to placebo across the two phase 3 studies. The Kaplan –Meier estimated Survival over time plots are shown for placebo and 10 mg/kg assigned groups in the placebo controlled phase. in this case the survival term means not experiencing a falling adverse event and the time is measured to the first fall, in case of multiple falling events. The placebo and 10 mg/kg groups separate around day 180 and the hazard ratio is 1.33 with an associated p-value of .016 suggesting a 33% relative increase in hazard of falling for 10 mg/kg. A quantitative integration of benefit and risk was not done but If the high dose increases falls it could be a significant risk for the indicated Alzheimer's population. This adverse effect would be on top of the amyloid related imaging abnormality edema risk and associated adverse effects. Finally, Again, long term safety data is less than it would be if the trials had not been stopped for futility.

Slide 23

Since this BLA submission contain one positive and one failed study with p=0.8 substantial evidence is obviously a critical question. A draft substantial evidence guidance was released in 2019. Some relevant excerpts are shown on this slide.

In particular the draft guidance mentions that common unblinding due to an effect of the drug, a primary endpoint that is subject to some bias when drug assignment is known could undermine an otherwise adequate and well controlled trial. Finally it states that if there are inconsistent trials there needs to be a collective assessment of all the trials and it may weaken the overall strength of evidence. Thus the draft guidance seems to provide some serious challenges to the evidence available in this application.

Slide 24

In the past couple years in the scientific community the issue of reproducibility of experiments has been raised . In this case we do not have a single strong study in isolation. On the contrary we actually have a second trial in which the purported effective dose was in the wrong direction compared to placebo, i.e., numerically worse than placebo. Under the null if winning in just one study out of two was enough then the chance of falsely rejecting the null would be .0975 across the two studies. Furthermore, if we select only the better study our estimate is very likely biased, and we already know not consistently repeatable in our experience. Thus excluding data from a large trial without sufficient justification is unscientific, statistically inappropriate and misleading.

Slide 25

To summarize, at best the evidence in this application is from study 302 only and there exists compellingly conflicting phase 3 evidence. If this NME were approved it would likely present challenges to other Alzheimer's candidates in development. For example, noninferiority would be questionable against these mixed divergent phase 3 results. Outside of noninferiority The need to demonstrate an add on effect to this drug could create possibly powering and or practical issues. Approval of this would also create recruitment and retention challenges for ongoing trials for other candidates which could potentially have more consistent effects.

Backup Slides