De Novo Request DEN160053 for the Neuronix Ltd.'s neuroAD Therapy System

Neurological Devices Advisory Panel Meeting
March 21st, 2019

FDA Presentations By: Erin Keegan, M.S., Claudette Brooks, M.D. Laura Thompson, Ph.D.,

Division of Neurological and Physical Medicine Devices (DNPMD)
Office of Device Evaluation (ODE)
Center for Devices and Radiological Health (CDRH)
Food and Drug Administration

neuroAD De Novo Review Team Members

Erin Keegan, MS	Lead/Engineering
Claudette Brooks, MD	Clinical
Peter Como, PhD	
Laura Thompson, PhD	Statistics
Jeffrey Silberberg, PhD	Electromagnetic Compatibility
Joseph Jorgens III, PhD	Software

Scope of Meeting

- Panel discussion on the specific questions
- Questions are for discussion only, no voting request
- Panel discussion will be incorporated into final review decision for neuroAD

Panel Questions

The Panel will be asked to discuss and make recommendations on:

1) Whether the U.S. pivotal study demonstrates a clinically meaningful benefit for the neuroAD as an adjunctive therapy.

2) When the neuroAD is used as an adjunctive therapy, the Panel will be asked to discuss and make recommendations on what minimum amount of improvement in ADAS-Cog alone is clinically meaningful, as well as the minimum amount of clinically meaningful improvement in the CGIC.

3) Whether the ADAS-Cog≤30 population is a clinically plausible subset and can patients be screened using the ADAS-Cog for the neuroAD

Panel Questions (continued)

4) Whether the post-hoc identification of the ADAS-Cog≤30 population at a later time point when no treatment is given is an adequate analysis of the pivotal study data, in concert with the supplemental data provided, to demonstrate probable benefit.

5) Are the risks for the neuroAD adequately reported and characterized.

6) Whether the probable benefits to health outweigh the probable risks.

Outline of FDA Presentations

- **Background Information**: Ms. Erin Keegan
- **Clinical Evidence**: Dr. Claudette Brooks
- **Statistical Considerations**: Dr. Laura Thompson
- **Benefit-Risk Considerations**: Dr. Claudette Brooks

Panel Q&A to FDA

Neuronix Ltd.'s neuroAD Therapy System: Background Information

Neurological Devices Advisory Panel Meeting
March 21st, 2019

Erin Keegan, M.S. – Biomedical Engineer and Lead Reviewer
Division of Neurological and Physical Medicine Devices (DNPMD)
Office of Device Evaluation (ODE)
Center for Devices and Radiological Health (CDRH)
Food and Drug Administration

Outline: Background Information

- De Novo Pathway
- Clinical Background of Alzheimer's Dementia
- neuroAD Device Description

Regulatory Background

- De Novo eligibility – neuroAD is eligible for evaluation in de novo because it:
  - Does not fit into any existing regulation (any device class)
  - Presents a moderate-risk profile

- If granted, FDA would likely place neuroAD in Class II (instead of Class I) and it would become a predicate device for the 510(k) pathway

- Pre/post-market data collection
  - Post-market studies are not intended to address pre-market questions
  - Benefits and risks need to be sufficiently characterized prior to market clearance

- Guidance: "Factors to Consider When Making Benefit-Risk Determinations in Medical Device Premarket Approval and De Novo Classifications"*

* https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm517504.pdf

Alzheimer's Dementia

- FDA is committed to ensuring medical devices have a reasonable assurance of safety and effectiveness for Alzheimer's dementia
- Alzheimer's Dementia (AD) is the most common dementia in the United States (US) and worldwide
- Progressive neurodegenerative disorder that impairs memory, thinking, language and behavior
- Neuropsychiatric symptoms: apathy, disengagement, or irritability
- Behavioral symptoms: aggression, wandering and various psychotic manifestations (hallucinations, delusions, and misidentification/misperception)

Approved AD Interventions

Drug Name	Brand Name	Approved Stage	Approval Year
Donepezil	Aricept	All stages	1996
Rivistagimine	Exelon	All stages	2000
Galantamine	Razadyne	Mild to moderate	2001
Memantine	Namenda	Moderate to severe	2003
Donepezil and Memantine	Namzaric	Moderate to severe	2014

There are no devices intended to treat Alzheimer's dementia that are approved for use in the United States

CAUTION: Should not use ADAS-Cog of Approved Drugs to inform MCID of ADAS-Cog Scale

Approved Treatments	neuroAD
• Cognitive and functional (or global) co-primary endpoints • Statistically significant difference between drug and placebo on both endpoints (approvals do not rely on a MCID) • Results replicated in at least two adequate and well controlled clinical trials comprising at least hundreds of subjects	• ADAS-Cog was sole primary endpoint • Sham outperformed active treatment at pivotal study primary endpoint. • One pivotal study and a collection of smaller studies with sample sizes <30

Drug approvals were not based on the changes in ADAS-Cog

Changes in ADAS-Cog on approved drugs do not inform MCID on ADAS-Cog Scale

neuroAD Regulatory Interactions

2010

First interactions, via pre-submissions

Topics included technical aspects of neuroAD, proposed nonclinical and prior clinical testing to support the proposed U.S. clinical study, and the study design of the proposed U.S. clinical study

2014

Pre-submission for the pivotal study

Determined to be not significant risk (NSR). Included updates to pivotal study design based on FDA feedback. It was determined de novo was appropriate pathway.

2016

The neuroAD system was granted Expedited Access Pathway (EAP) designation on May 3.

The Data Development Plan for the proposed EAP designation was approved.

Direct de novo submitted (DEN160053)

2017

Request for additional information sent March 13, 2017

Meeting with the sponsor to discuss deficiencies and proposed responses

Sponsor submits response to request for additional information

2018

Network of Experts interactions

Denial letter issued June 22, 2018.

Sponsor appeals denial decision

FDA reopens the submission in response to appeal

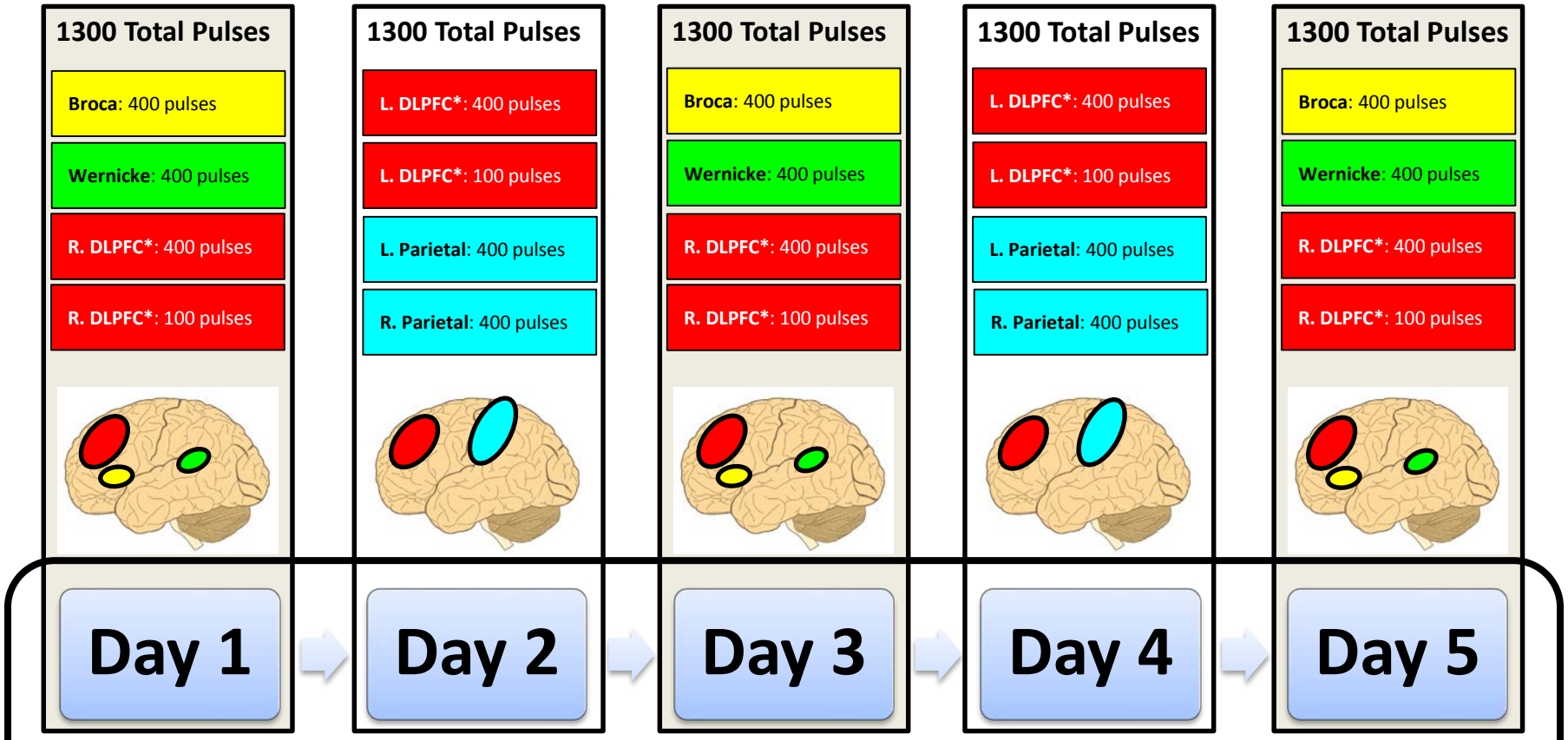
Indications for Use (As Proposed by Neuronix)

- The neuroAD™ Therapy System is intended for neuro-stimulation concurrently combined with cognitive training.
- The neuroAD™ Therapy System is indicated for the treatment of mild to moderate dementia of the Alzheimer's type in patients with a baseline ADAS-Cog score up to 30.
- The neuroAD™ Therapy System may be used in conjunction with other pharmacological and non-pharmacological therapies.

Device Description: Important Points

- TMS
  - Intensity was patient-specific, based on motor threshold (MT)
- Cognitive Training
  - Training difficulty progresses individually based on performance via algorithm
  - Impact of increasing/decreasing difficulty level not pre-specified for assessment in US pivotal study
- Treatment Paradigm

Treatment Paradigm



This set is repeated each week. After every week, the software calculates patient performance on the cognitive training modules and adjusts the difficulty of the training according to a pre-specified algorithm.

Week 1 → Week 2 → Week 3 → Week 4 → Week 5 → Week 6

Treatment Ends

(Primary endpoint of Pivotal Study at Week 7)

* DLPFC = dorsolateral prefrontal cortex. Note changes between left ("L.") and right ("R.") targets for DLPFC and parietal targets.

Clinical Testing: Important Points

- US Pivotal is primary dataset
  - Pre-specified results have most certainty
- Supplemental studies are small, high uncertainty
- Post-Hoc analyses provided by the sponsor
- Nominal p-values

Neuronix Ltd.'s neuroAD Therapy System: Clinical Evidence

Neurological Devices Advisory Panel Meeting
March 21st, 2019

Claudette Brooks, MD – Neurologist and Clinical Reviewer
Division of Neurological and Physical Medicine Devices (DNPMD)
Office of Device Evaluation (ODE)
Center for Devices and Radiological Health (CDRH)
Food and Drug Administration

Outline: Clinical Evidence

- Pivotal Study Design
- Pivotal Study Results
- Pivotal Study Post-Hoc Subgroup Analysis
- Additional Post-Hoc Analyses
  - Korea Studies
  - Supplemental Investigations

Pivotal Study: Important Points

- The pivotal study did not meet its pre-specified primary endpoint (sham out-performed neuroAD)
- The safety data from the pivotal study are consistent with a moderate-risk profile device
- There were no significant concerns with the design and conduct of the pivotal study in its prespecified form and statistical analysis plan

Pivotal Study Design

- Enrolled mild to moderate AD patients (baseline ADAS-Cog above 17)
- Two Groups (Cognitive training (CT) not studied separately):
  - Active (active TMS and active CT)
  - Sham (sham TMS and sham CT)
- Active or Sham Treatment lasted 6 weeks, 5 days/week (in clinic)
- Studied as an adjunctive therapy – AD medications were required to be stable for at least 60 days prior to joining the study, and were monitored throughout

US Pivotal Study Endpoints

- Primary Effectiveness Endpoint:
  - Change in ADAS-Cog from baseline to 7 Weeks
- Secondary Endpoints:
  - Change in ADAS-Cog from baseline to 12 Weeks (to assess durability)
  - Change in ADCS-CGI-C from baseline to 7 Weeks
  - Change in ADCS-CGI-C from baseline to 12 weeks
- Primary Safety Endpoint:
  - All AEs occurring at any point in the trial, regardless of relation to study device or procedure
- Endpoints assessed at 7 weeks were approximately 1-week post-treatment and endpoints assessed at 12 weeks were approximately 6 weeks post-treatment

Pivotal Study Effectiveness Assessment Scales

ADAS-Cog

- 70-point scale (higher score, greater dysfunction)
- ≤ five “normal”; ≥18 considered impaired
- Clinician-administered patient evaluation
- Measures cognitive domains including memory, language and praxis
- The test administrator adds up points for the errors in each task of the ADAS-Cog for a total score
- **Important for intervention: Negative change from baseline would suggest improvement on the scale**

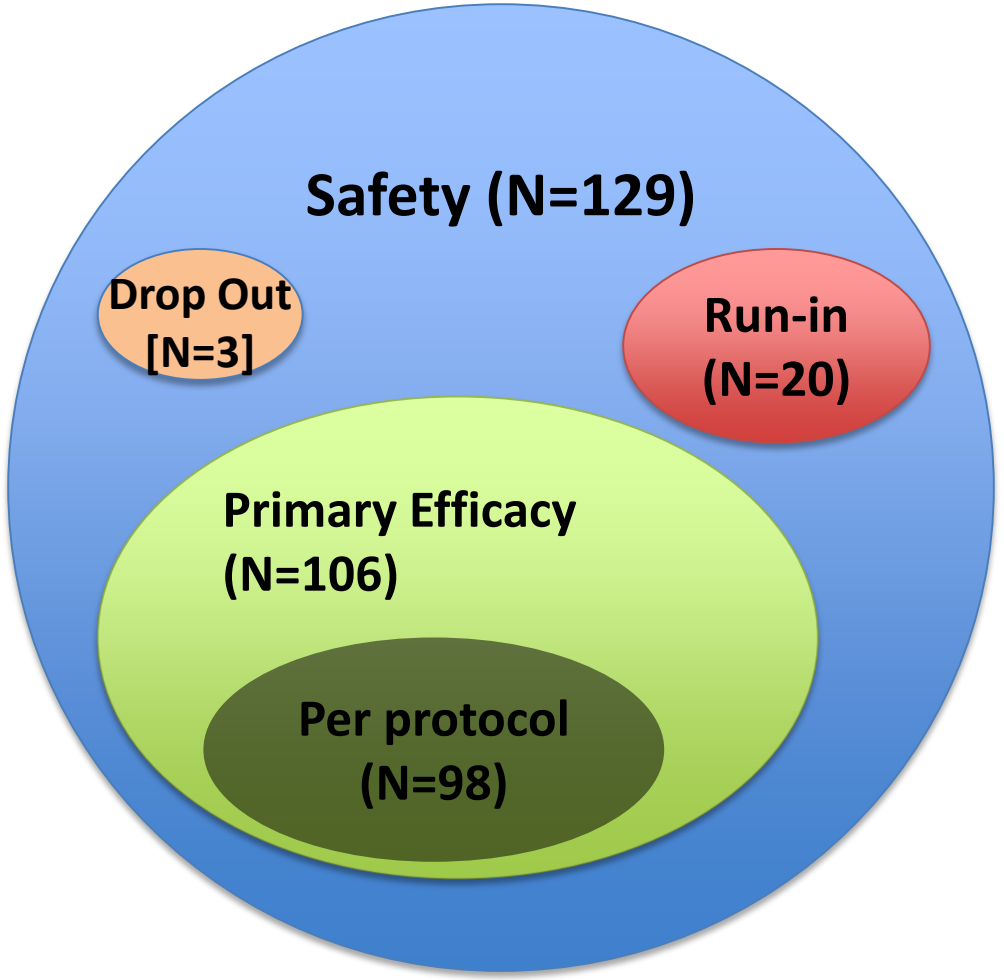
CGI-C

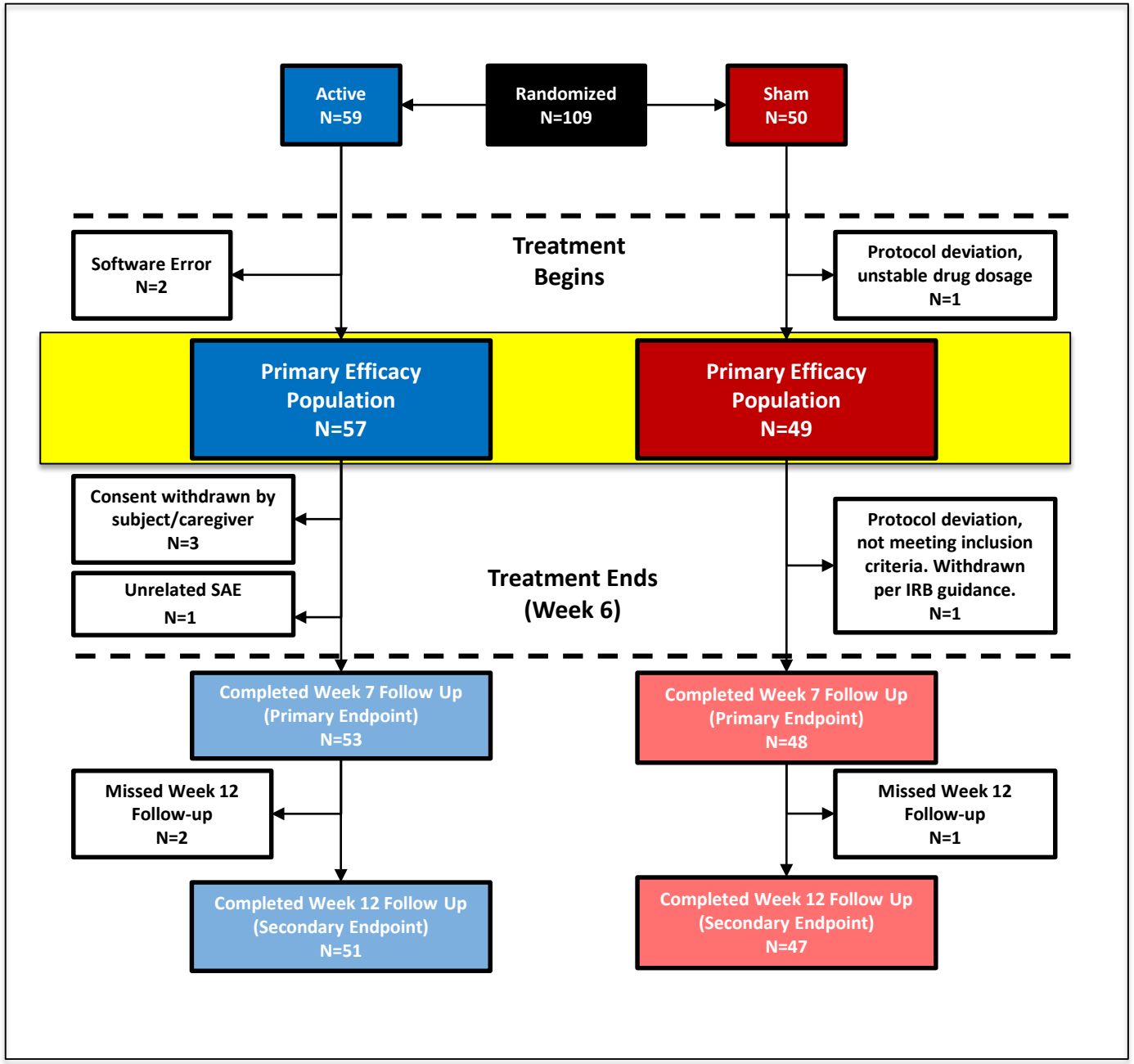
- 7-point scale (1=very much improved; 4=no change; 7=very much worse)
- Clinician rated, based on patient +/- informant interview
- Requires the assessor to consider a number of cognitive, functional, and behavioral areas prior to providing an overall “global” assessment of clinical change
- Worksheet lists relevant symptoms potentially useful in judging clinically meaningful change and allows for notes for future reference
- **Important for intervention: A score of ‘4’ would indicate no change**

Statistical Analysis Populations

- Statistical analysis populations
  - **Primary Safety**
  - **Primary Efficacy (PE)**
  - Per Protocol (PP) – not discussed in this presentation

- Run-in subjects
  - First two subjects at each site (n=20)
  - Not randomized, counted as “active”
  - Included in safety only

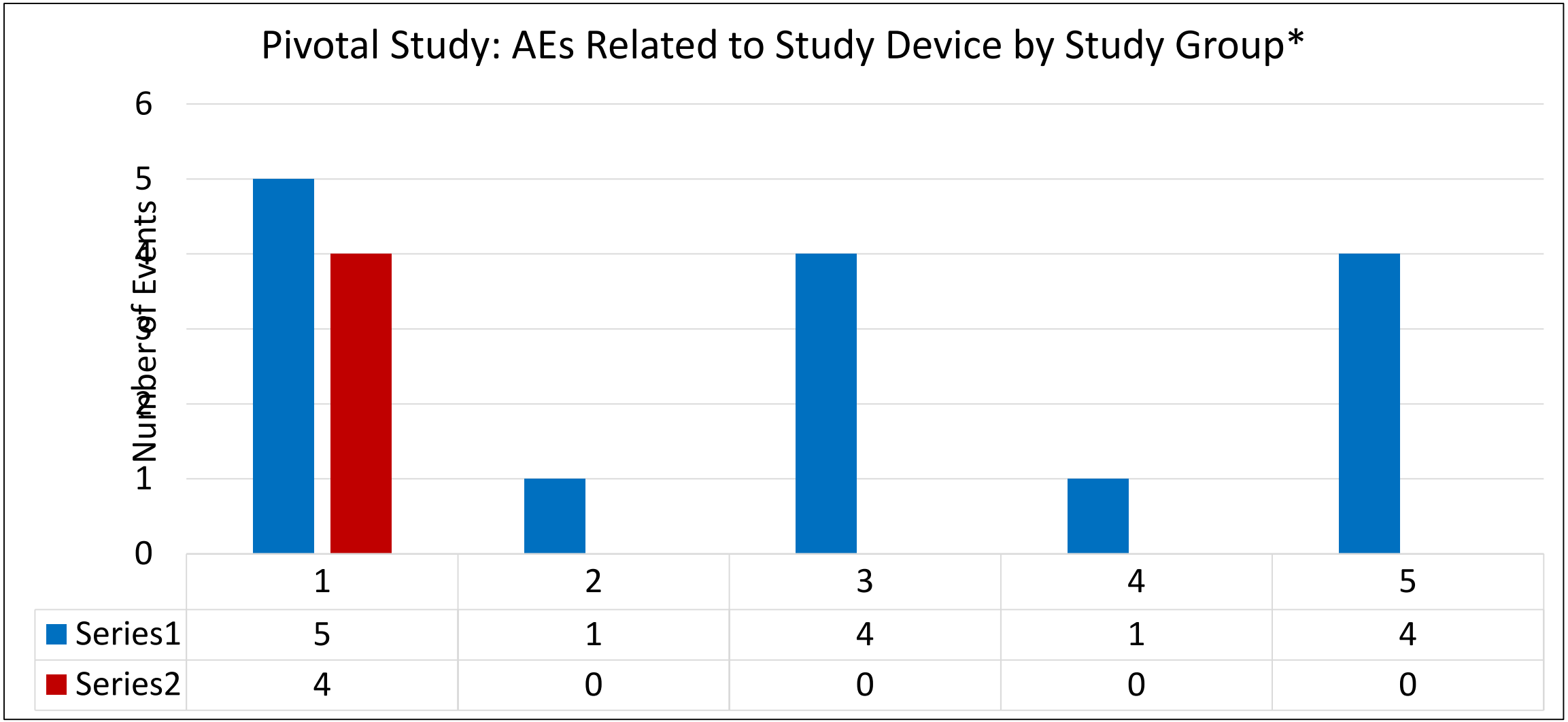




Patient Disposition

The Safety population (N=129) minus the Run-in subjects (N=20) yields the Randomized population (N=109)

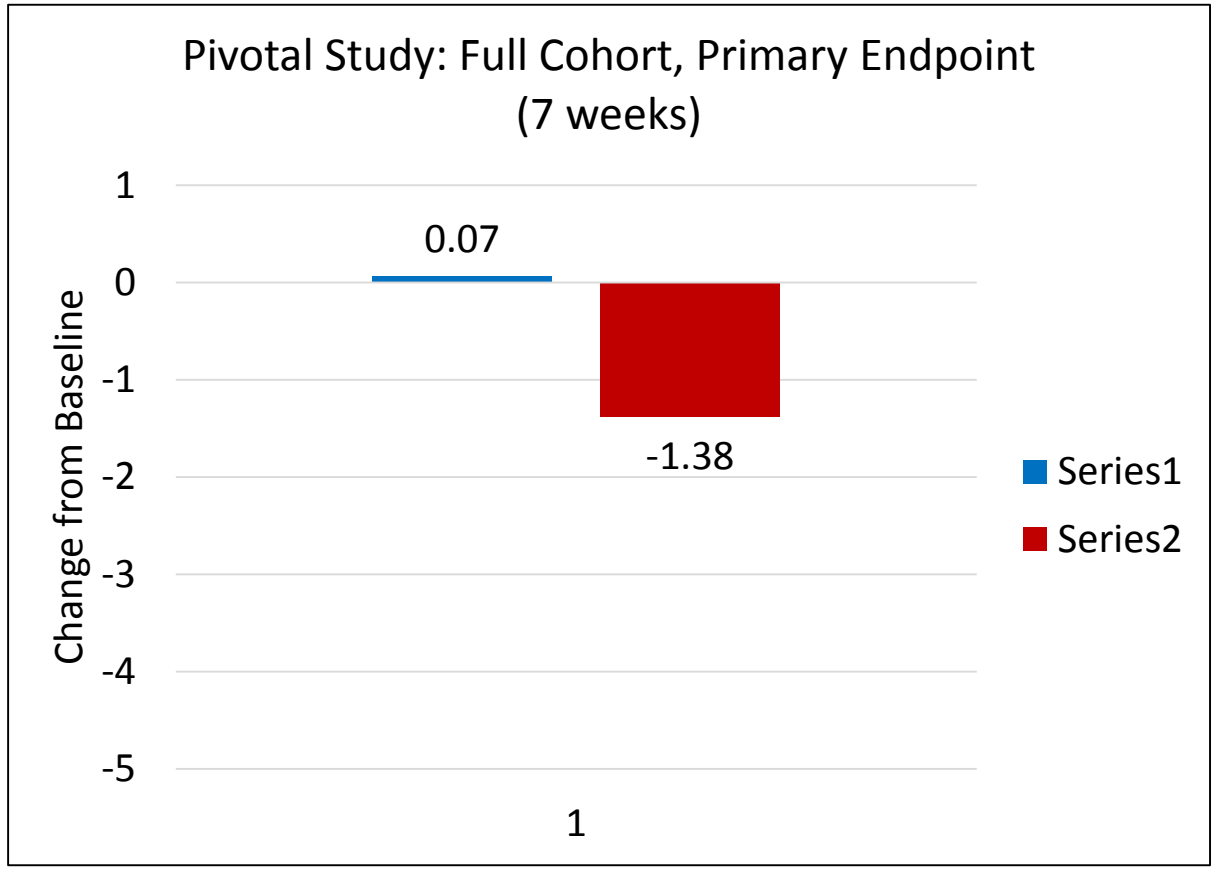
Primary Safety Endpoint Results (N=129)

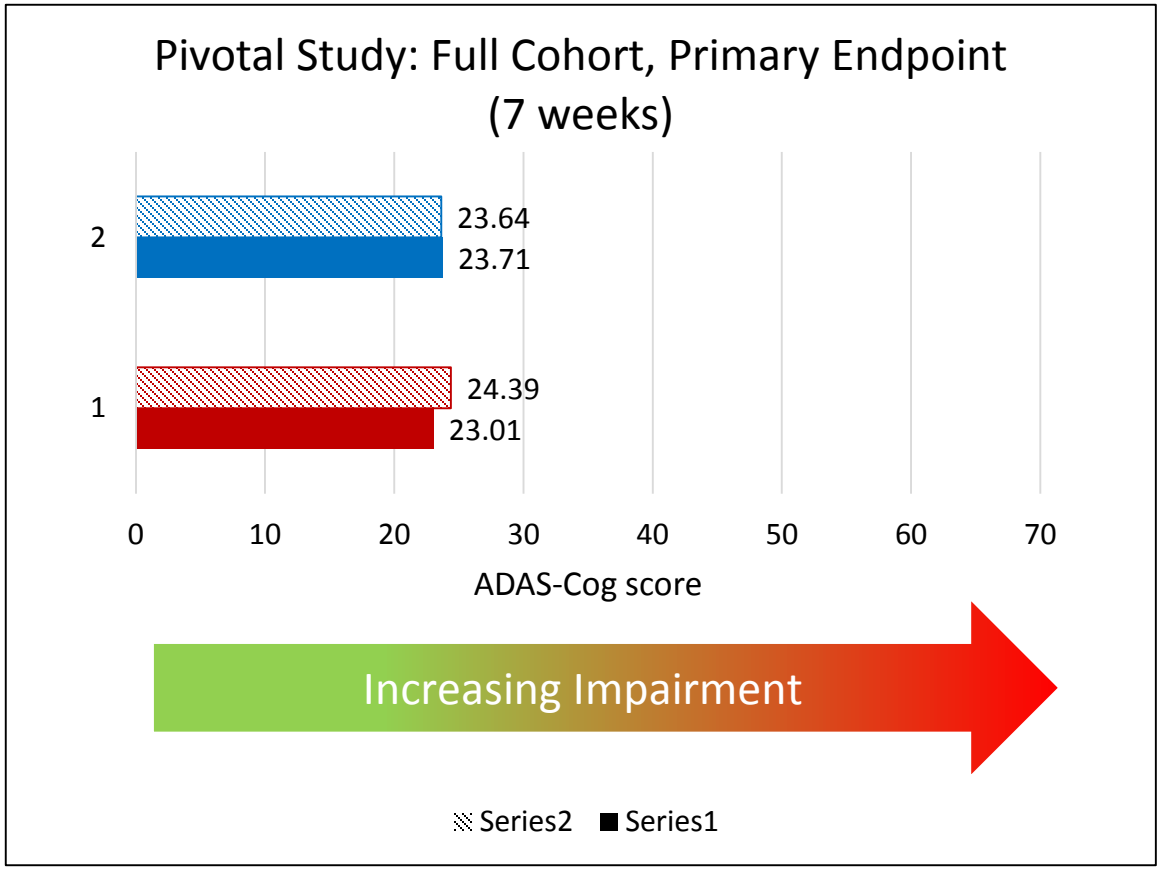


* Possible/Probable/Definite relationship. A subject may have experienced more than one type of event.

Primary Effectiveness Endpoint Results

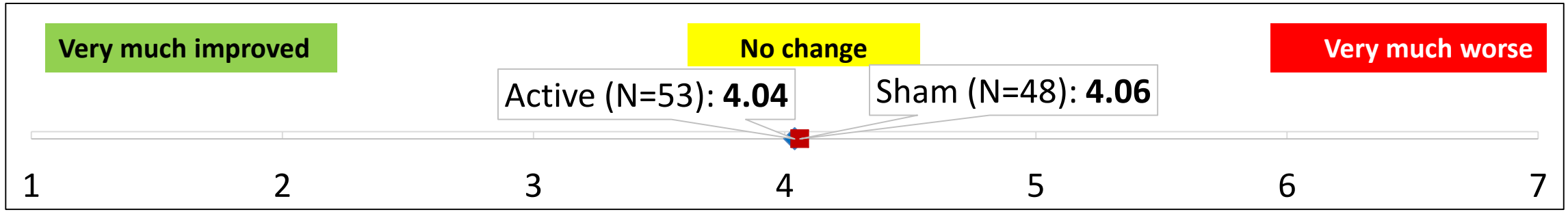
The difference between groups from Baseline to 7 weeks on ADAS-Cog was **1.45 points, in favor of sham** (p=0.09).



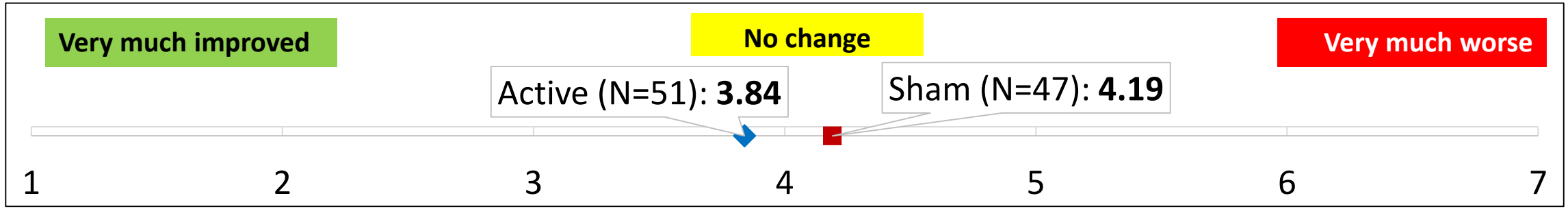


Secondary Effectiveness Endpoints Results: CGI-C

Difference from Baseline to 7 weeks on CGI-C between groups was 0.02 in favor of treatment (p=0.96)



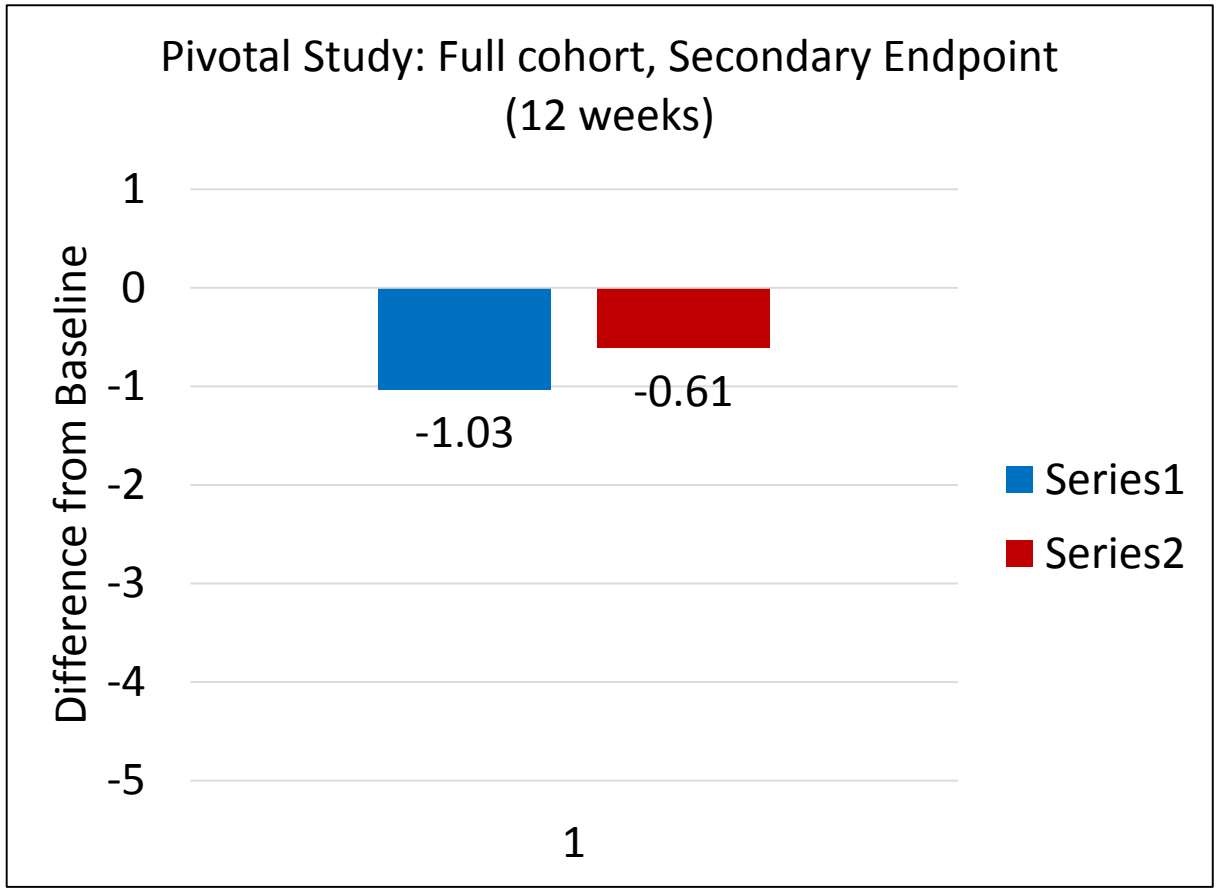
Difference from Baseline to 12 weeks* on CGI-C between groups was 0.35 in favor of treatment (p=0.12)

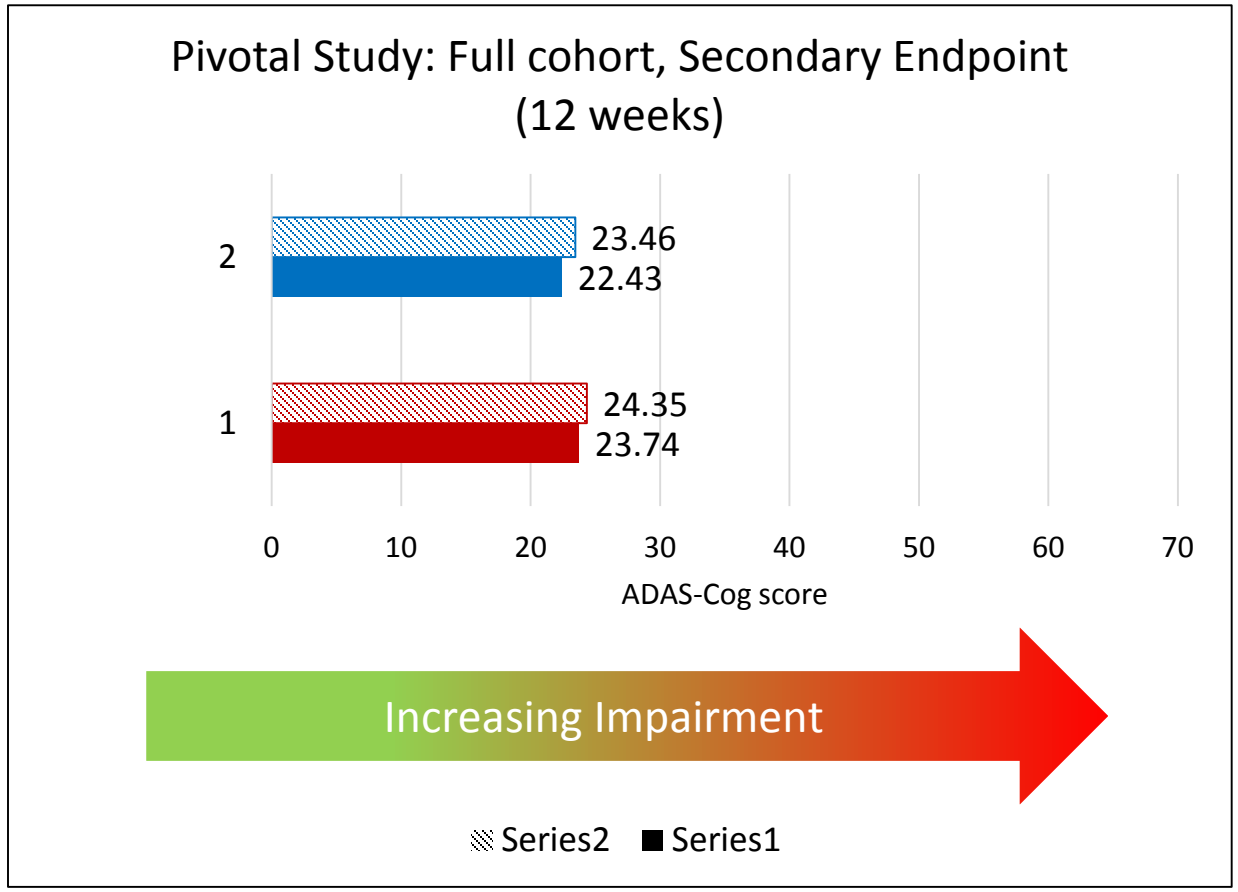


*There was no neuroAD treatment or sham intervention between 7 and 12 weeks

Secondary Effectiveness Endpoint Result: ADAS-Cog

Difference from Baseline to 12 weeks* on ADAS-Cog between groups was -0.42 in favor of treatment (p=0.64)





*There was no neuroAD treatment or sham intervention between 7 and 12 weeks

Summary: Pivotal Study, Pre-specified Results

- The study did not meet the primary effectiveness endpoint – sham outperformed active treatment on ADAS-Cog at 7 weeks compared to baseline

- Secondary endpoints only begin to favor the active group at the 12-week visit, six weeks after the end of treatment

- Cause of these differences is not clear

SPONSOR’S PIVOTAL STUDY POST-HOC ANALYSIS: BASELINE ADAS-COG ≤30

Pivotal Study: Post-hoc Analysis Population (Baseline ADAS-Cog ≤30)

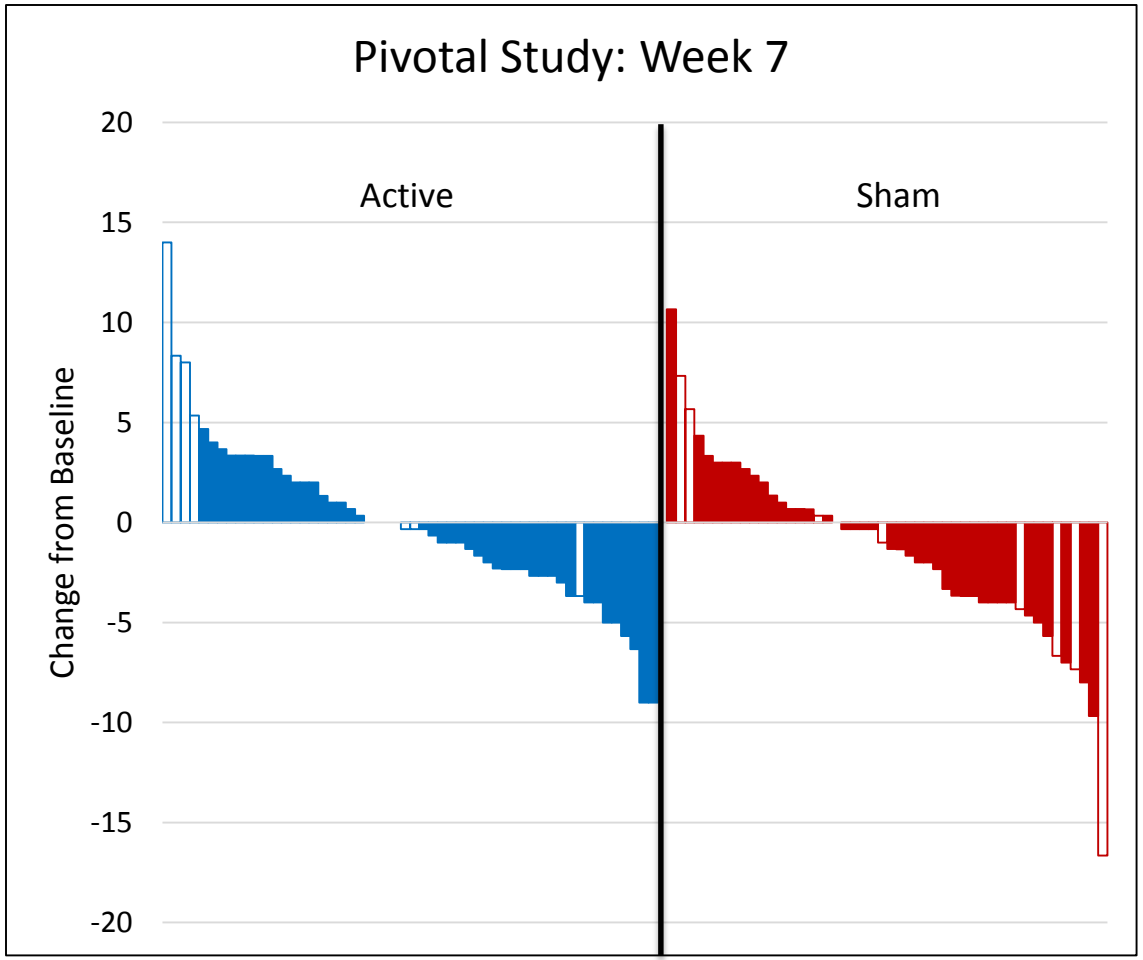
- FDA requests pre-specified analyses to limit bias and reduce uncertainty – there is a higher element of chance in any post-hoc result

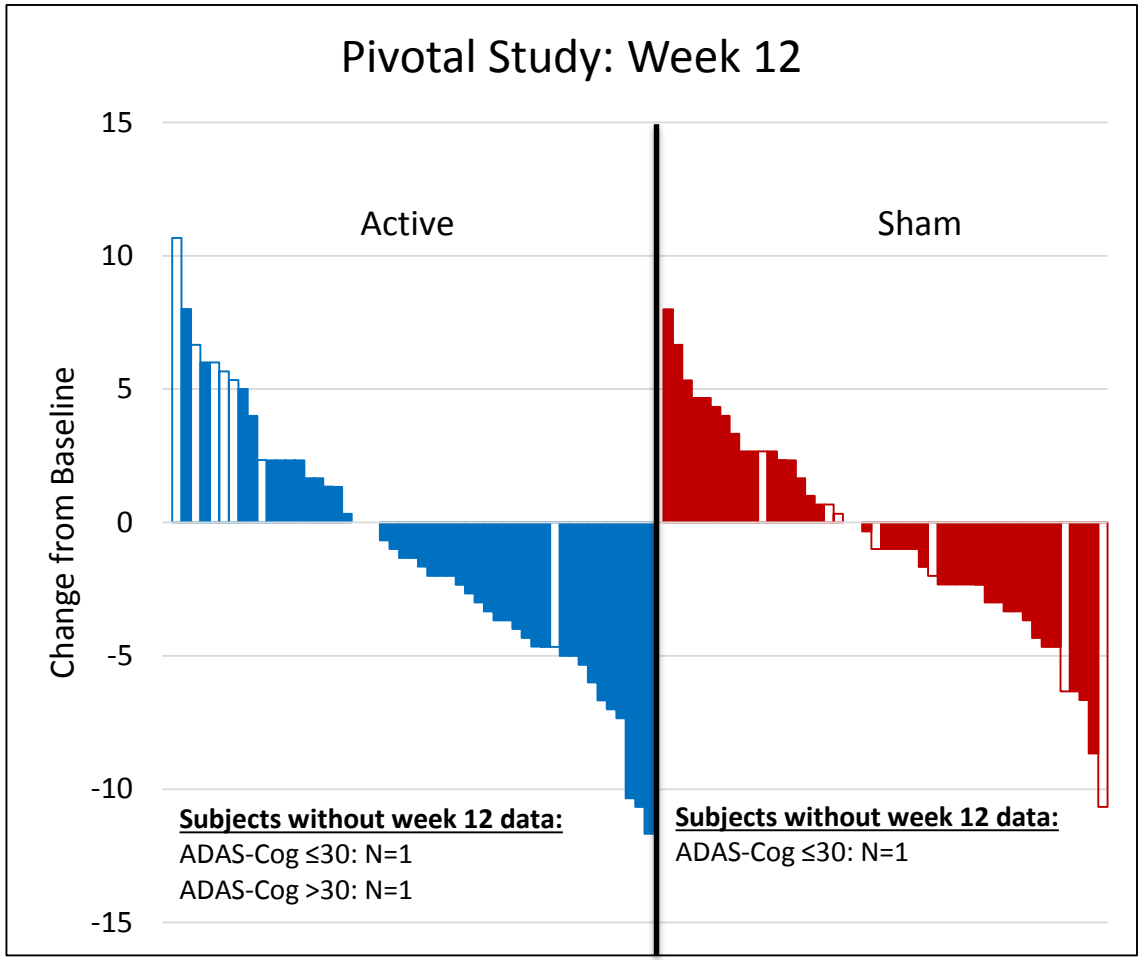
- FDA considers post-hoc analyses hypothesis-generating and recommends verification on independent data set

- Basis for sub-group selection and specific ADAS-Cog cut off point appears to be post hoc analysis of Pivotal study data

Patient-Level Data by Baseline ADAS-Cog

Exclusions tend to be poor-performing active subjects and high-performing sham subjects

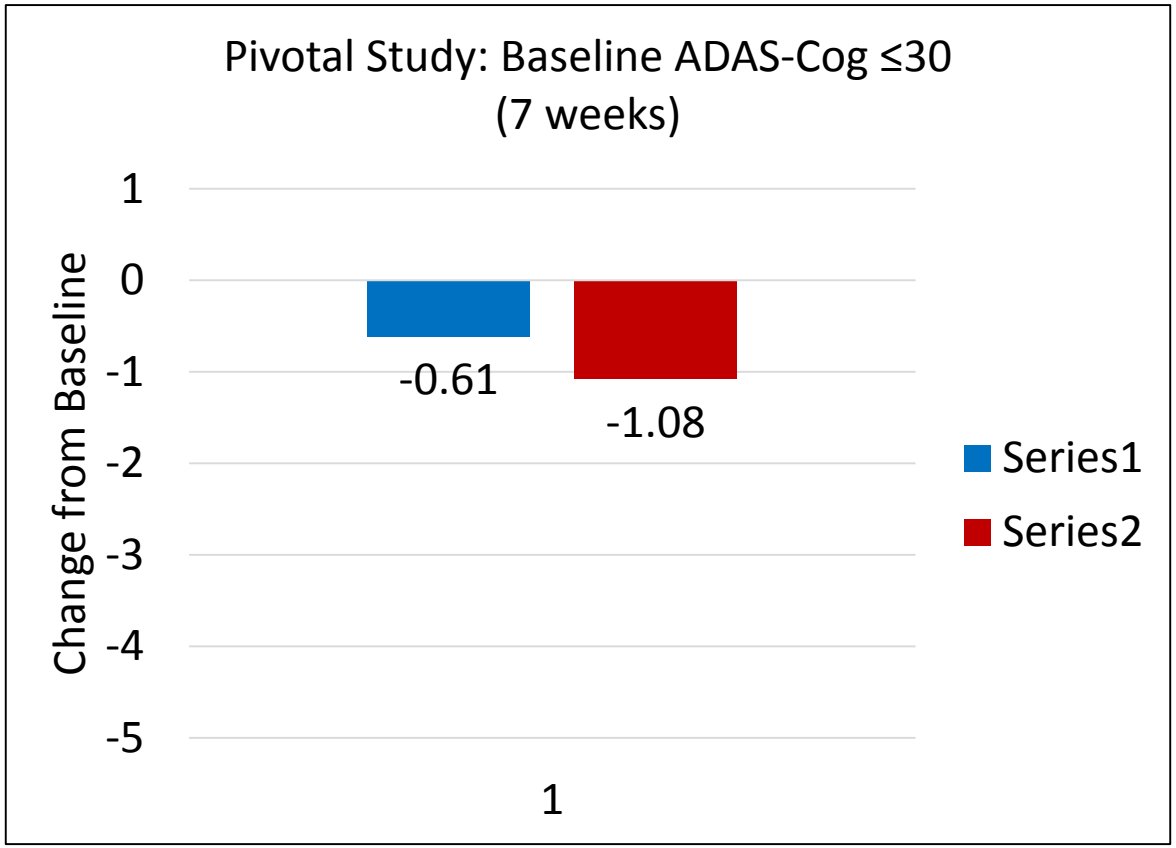


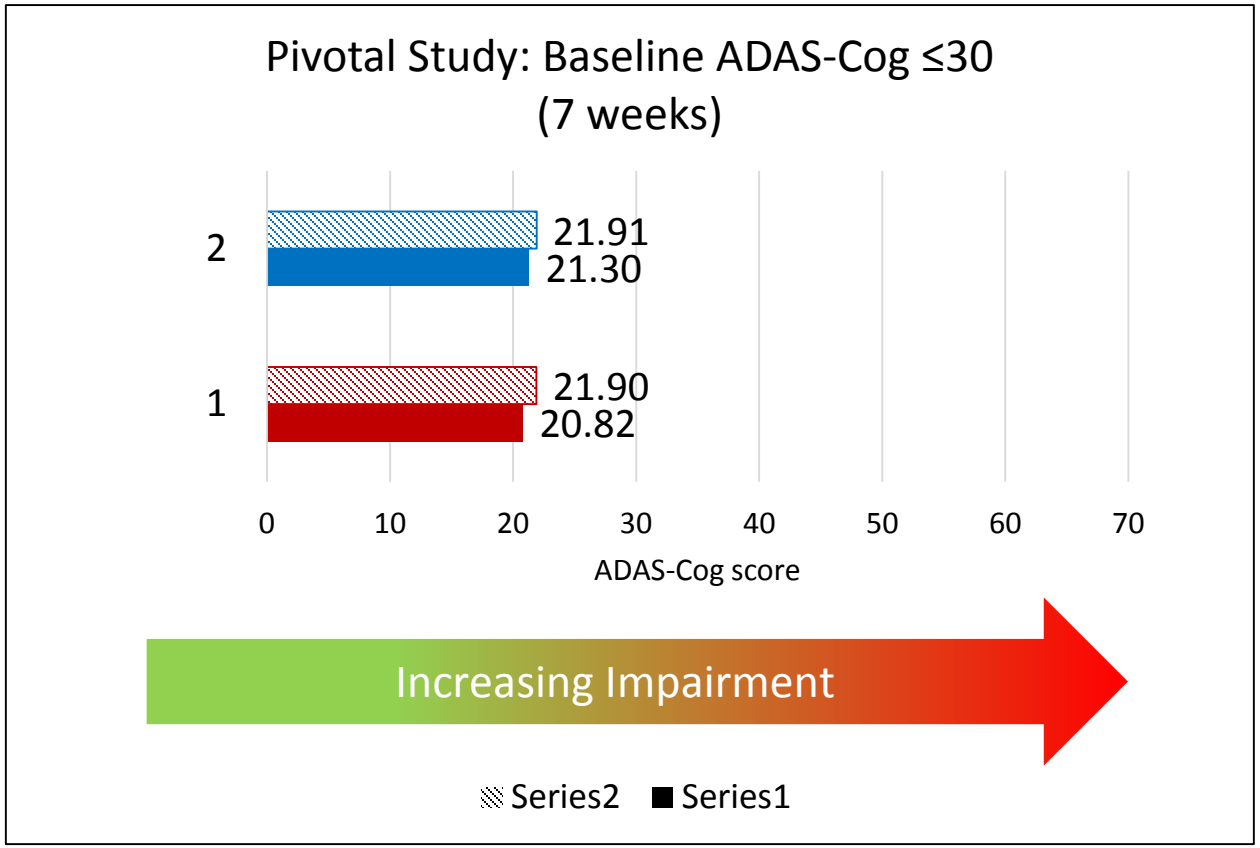


Each bar represents one subject. Filled bars indicate a subject with baseline ADAS-Cog ≤30 – empty bars indicate baseline ADAS-Cog >30.

Post-Hoc Subgroup, Baseline ADAS-Cog ≤30

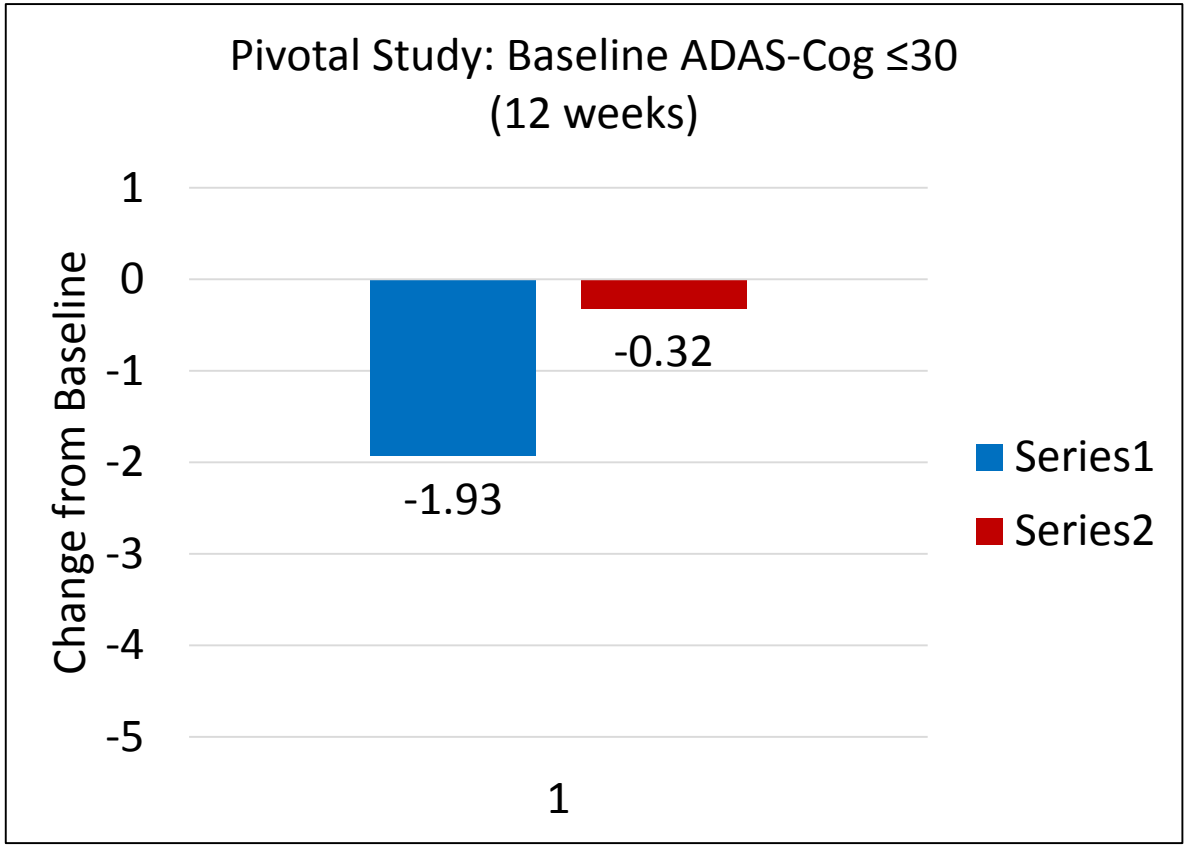
In the subgroup, the difference from Baseline to 7 weeks on ADAS-Cog between groups was **0.47 points, in favor of sham**

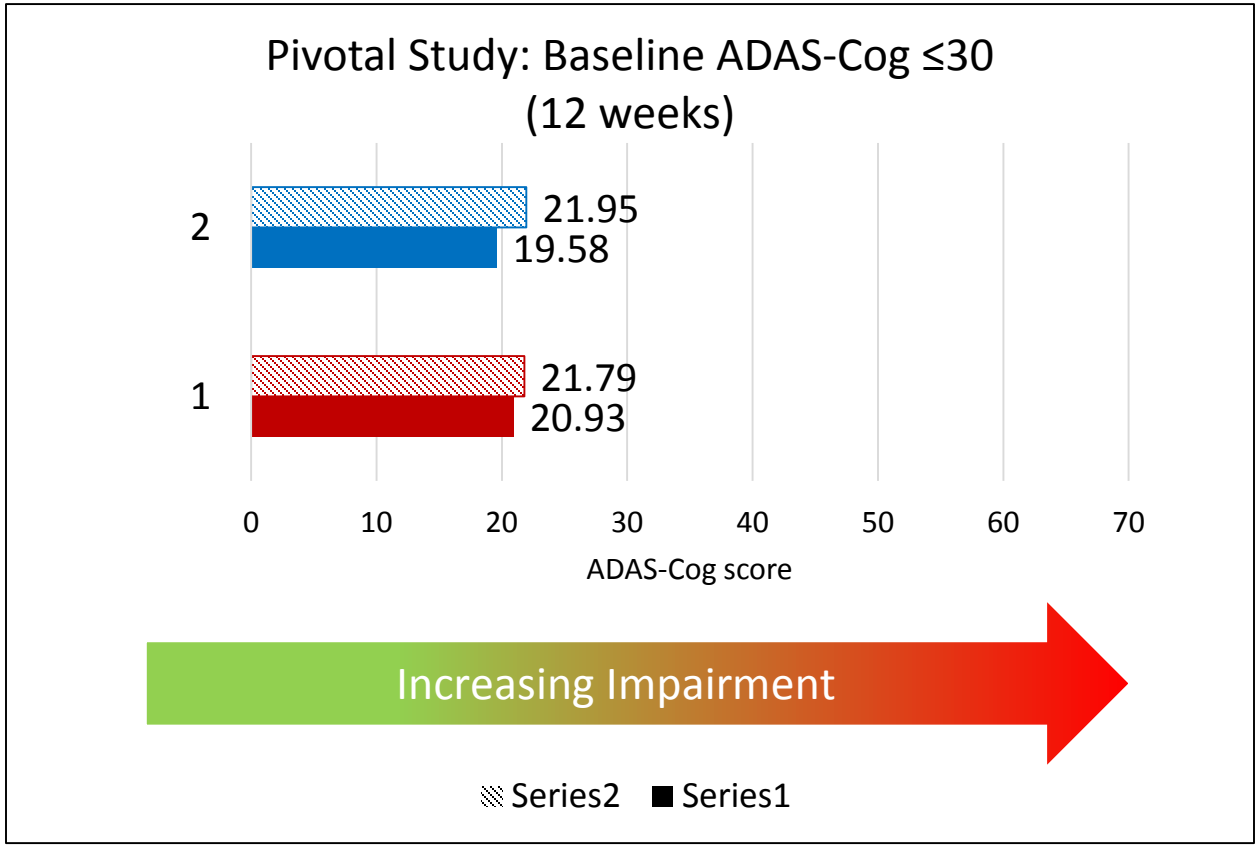




Post-Hoc Subgroup, Baseline ADAS-Cog ≤30

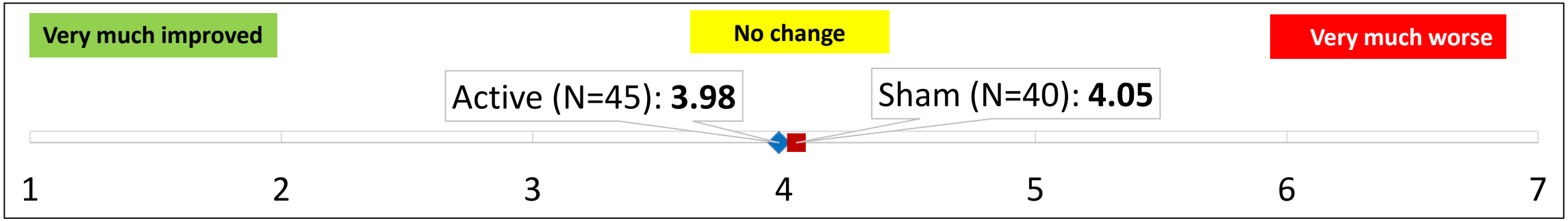
In the subgroup, the difference from Baseline to 12 weeks on ADAS-Cog between groups was **1.61 points, in favor of treatment**



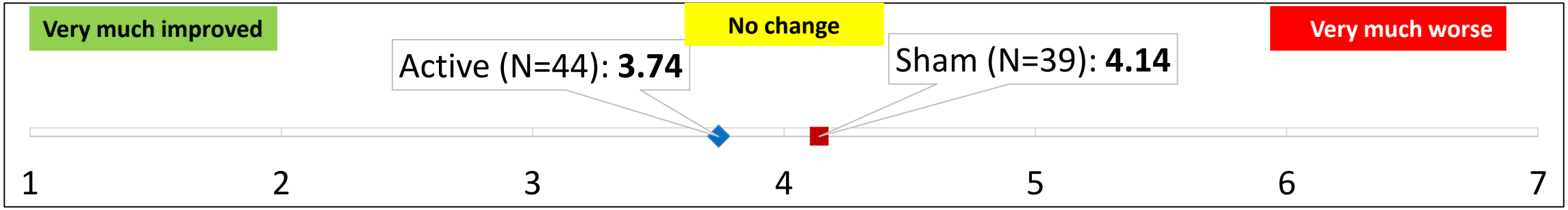


Post-Hoc Subgroup, Baseline ADAS-Cog ≤30

In subgroup, difference from Baseline to 7 weeks on CGI-C between groups was **0.07 in favor of treatment**



In subgroup, difference from Baseline to 12 weeks on CGI-C between groups was **0.40 in favor of treatment**



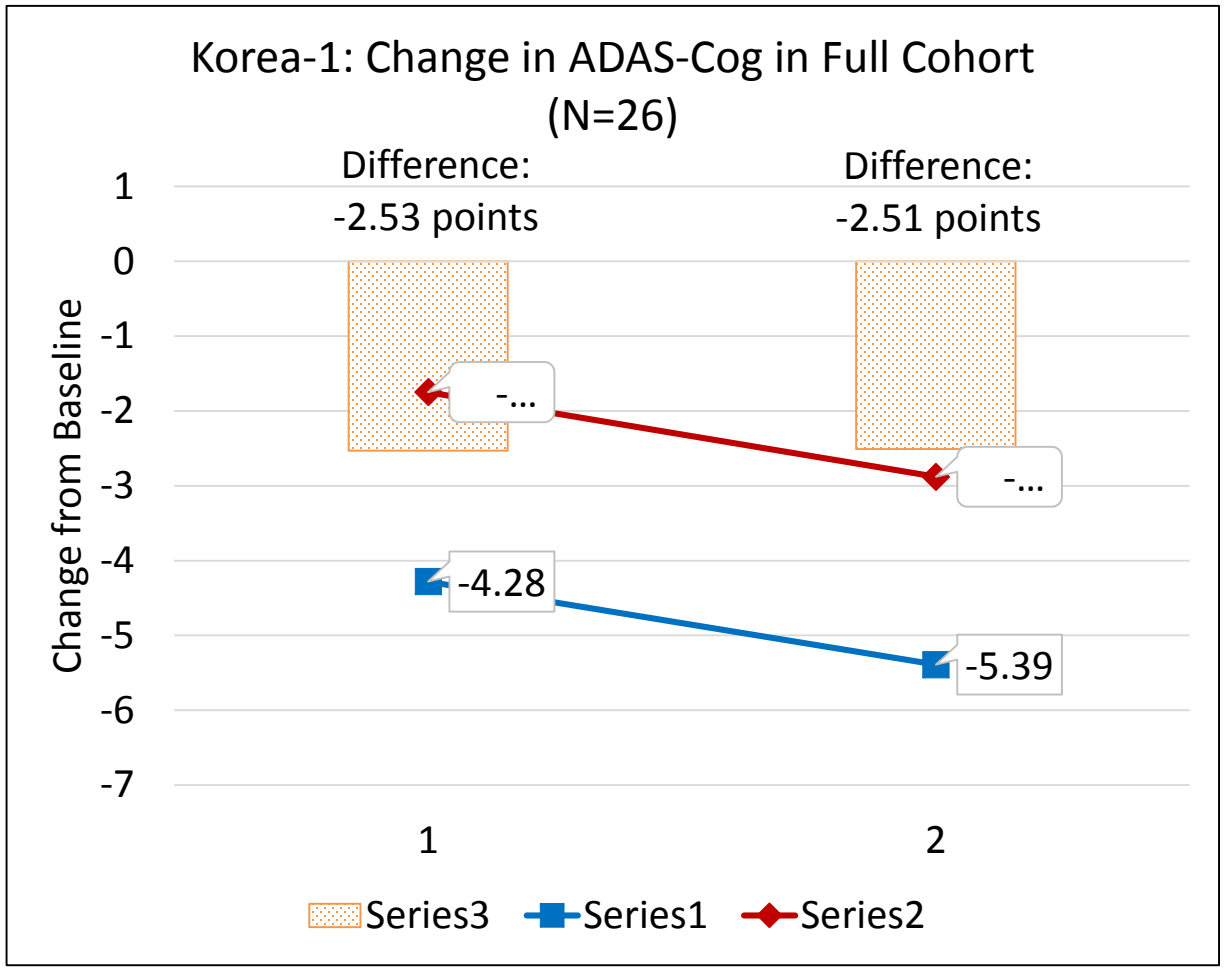
ADDITIONAL POST-HOC ANALYSES: KOREA STUDIES "1" AND "2"

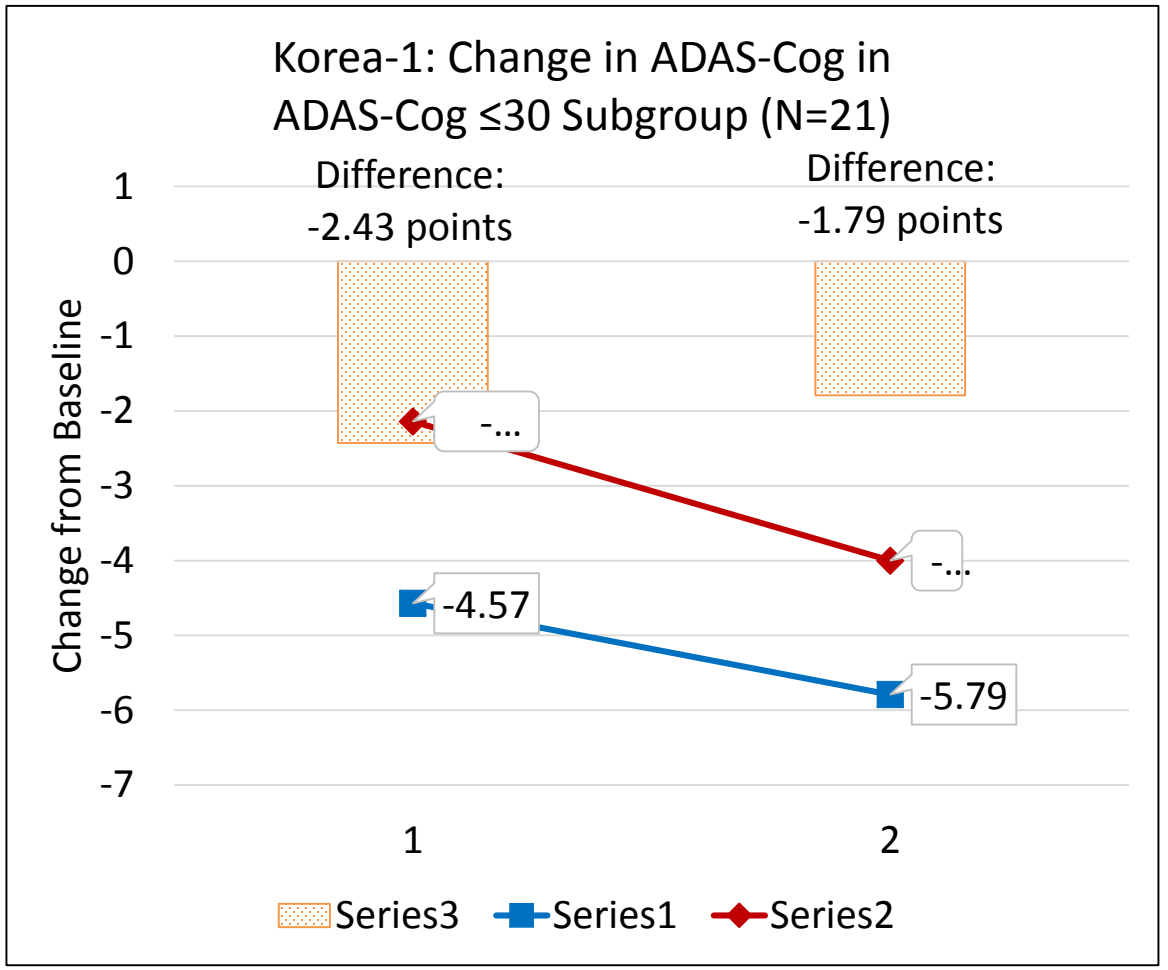
Overview of Korea Study Designs

- Similar protocols to US Pivotal
- Small sample sizes
- Korea-1 (“Pilot”, Lee et al., 2016)
  - 27 subjects; 18 active, 9 sham (data available from 26 subjects)
  - Not limited to Baseline ADAS-Cog ≤30
- Korea-2 (“Pivotal”)
  - 22 subjects; 11 active, 11 sham
  - Interim analysis; study enrollment currently suspended pending FDA decision

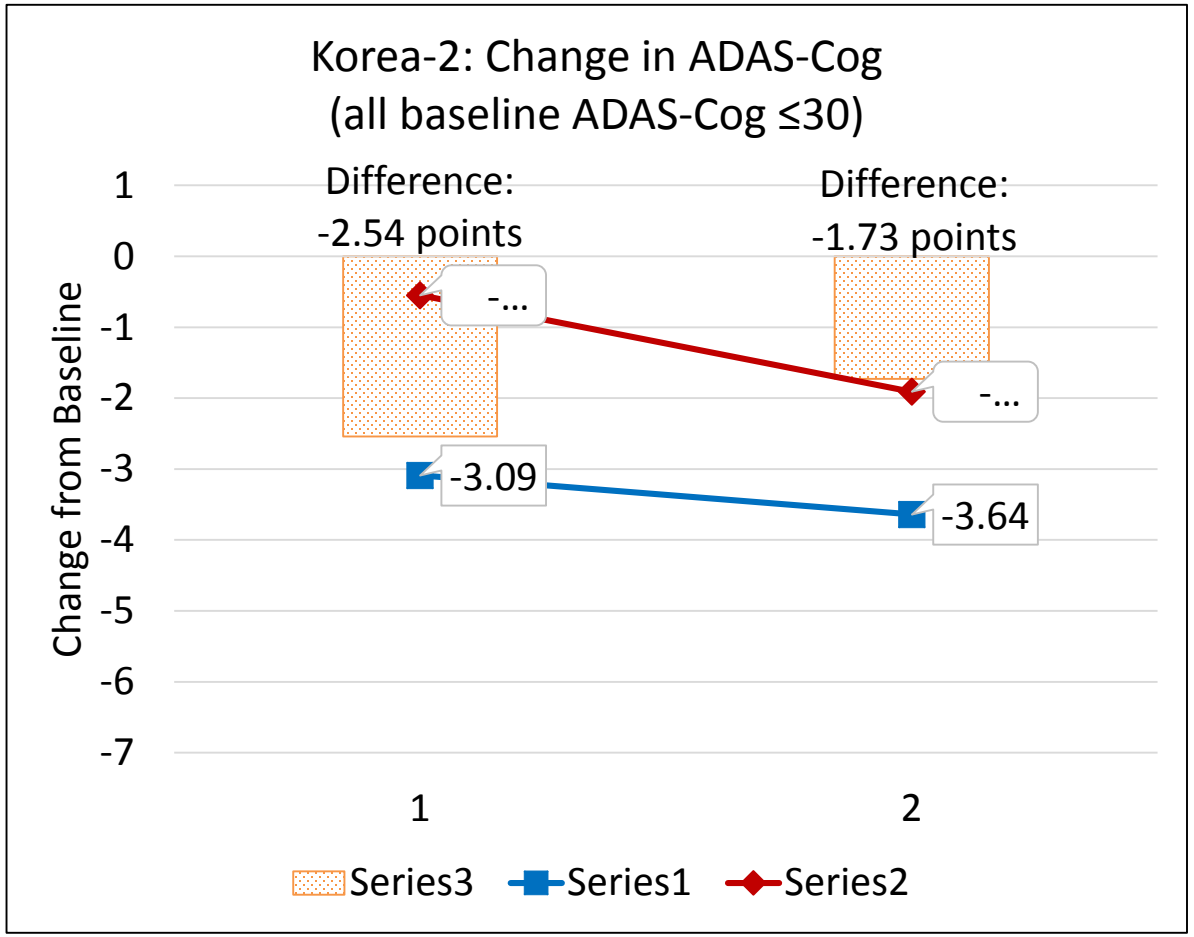
Korea-1 Effectiveness Results by Subgroup

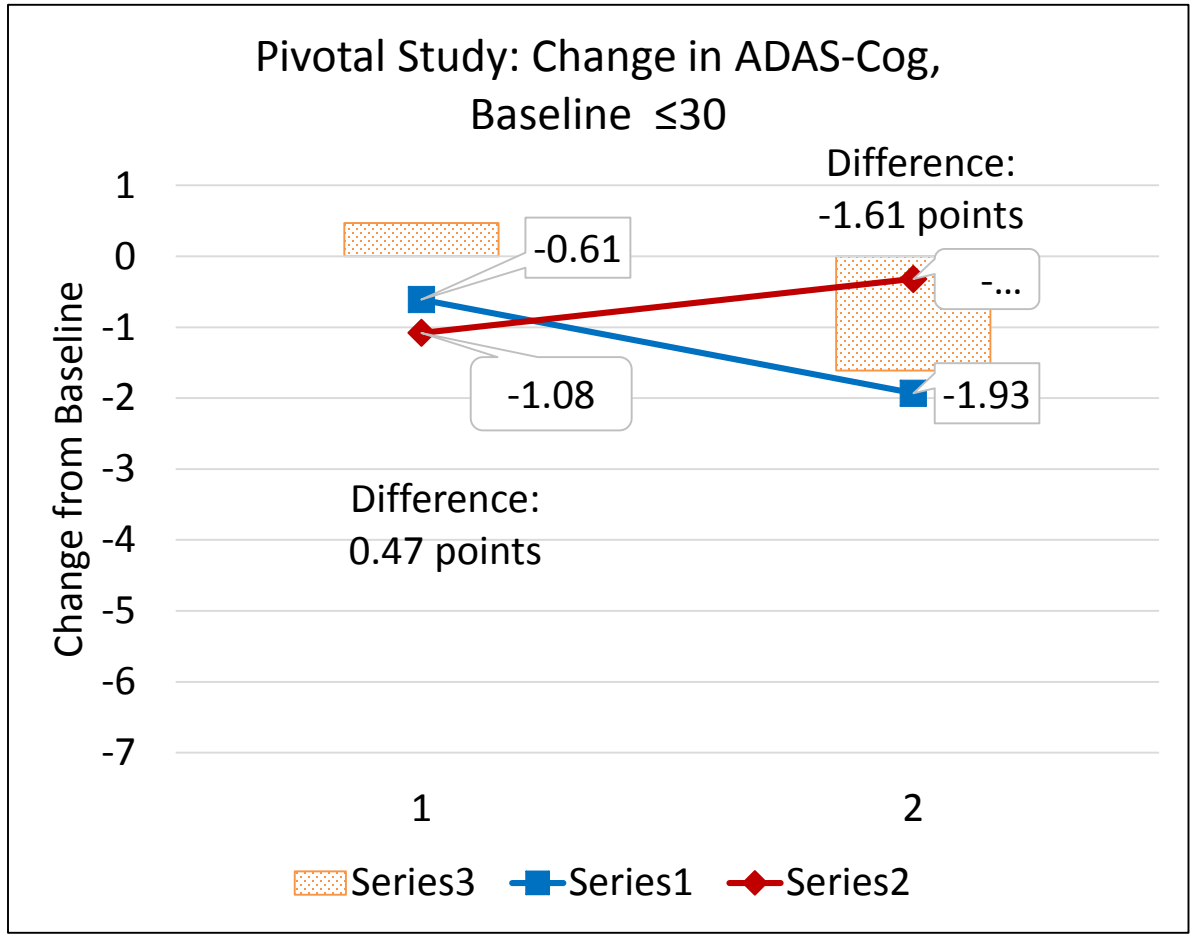
Both active and sham groups show similar trends in improvement over time, both for the full cohort and subpopulation



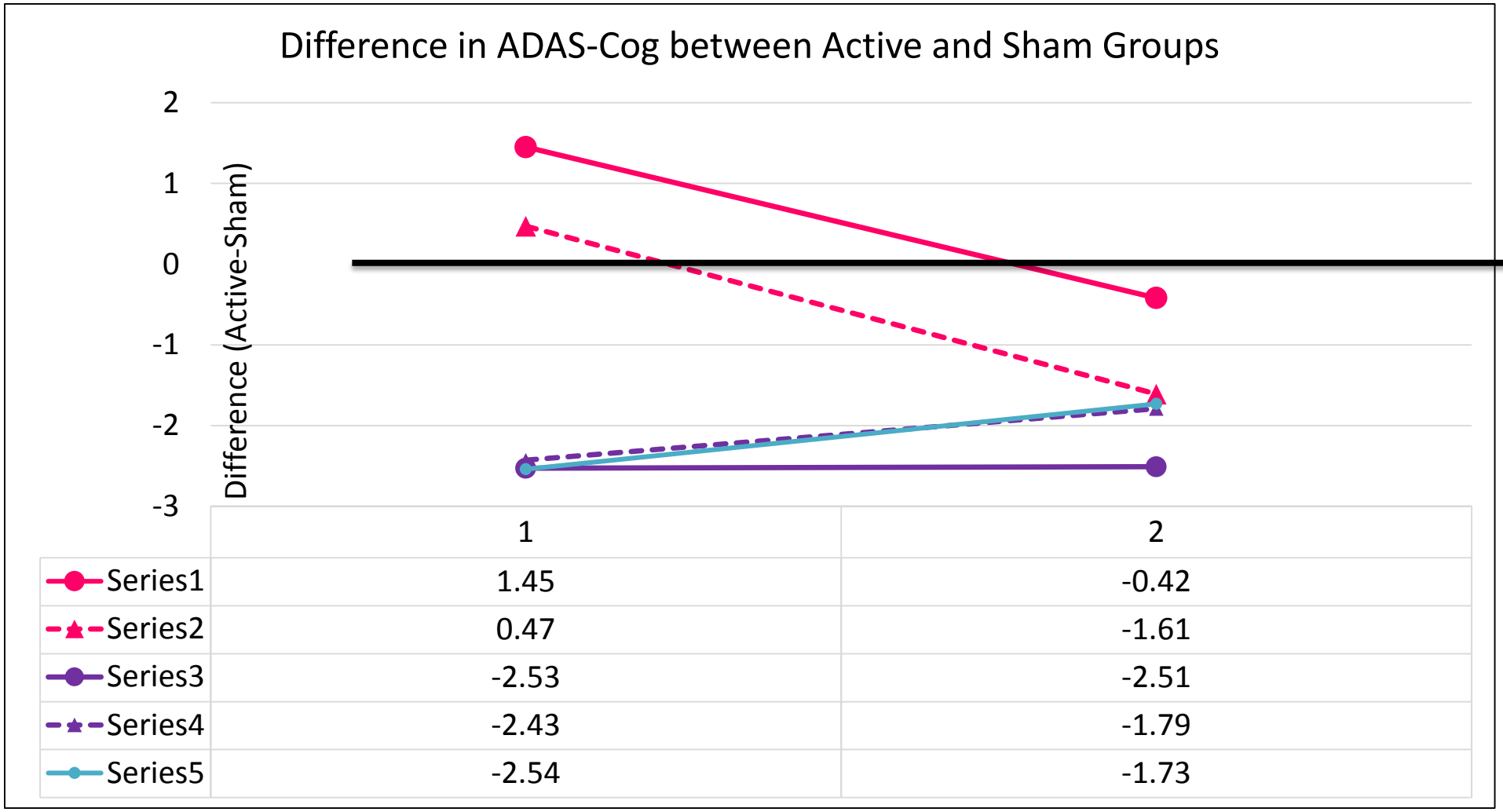


Performance of ADAS-Cog ≤30 Subgroup Between Korea-2 (interim) and Pivotal Studies





Mean Difference in ADAS-Cog Over Time Between Studies

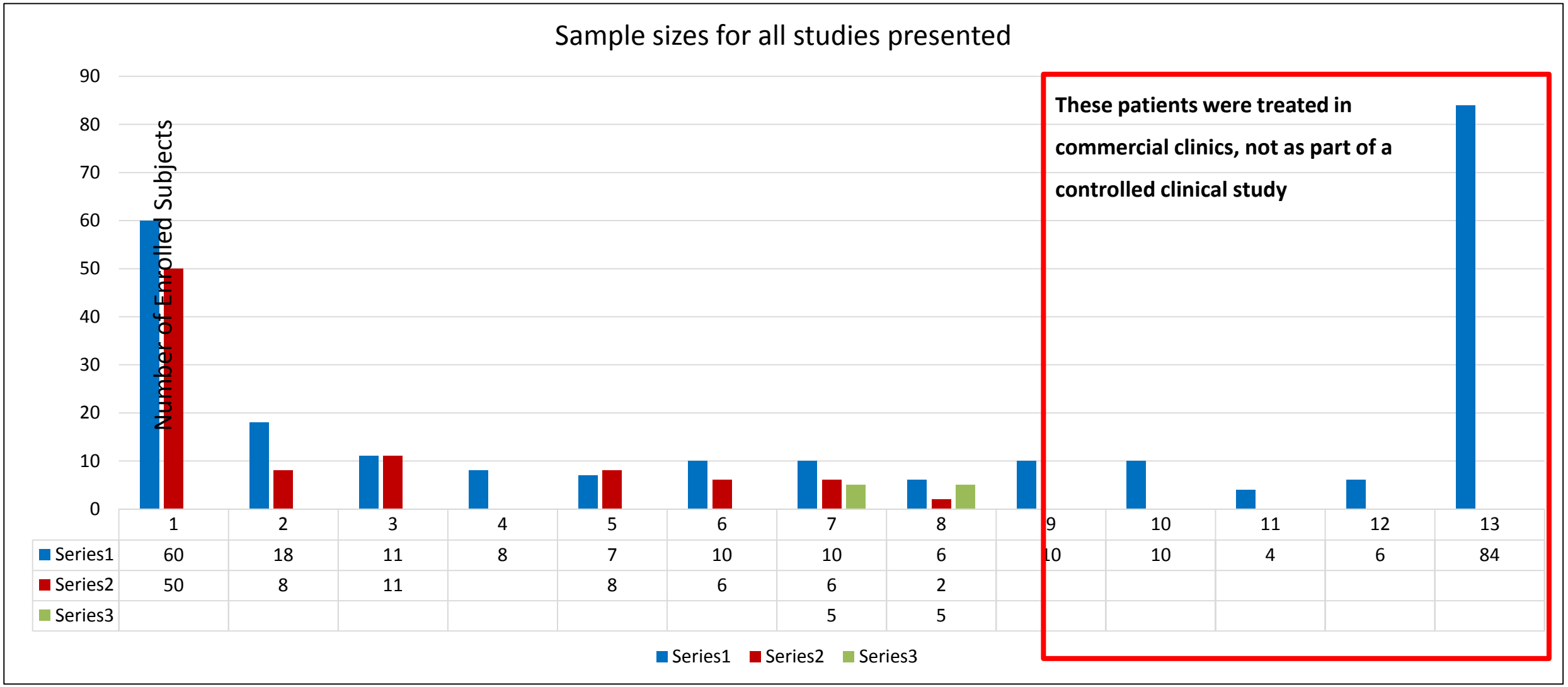


Positive between-group differences favor the sham group

Negative between-group differences favor the treatment group

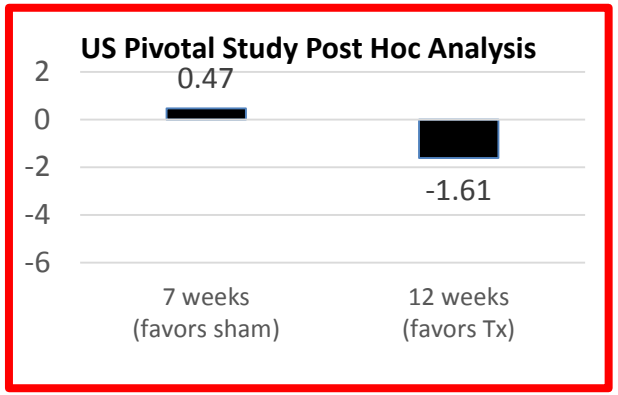
ADDITIONAL POST-HOC ANALYSES SUPPLEMENTAL INVESTIGATIONS

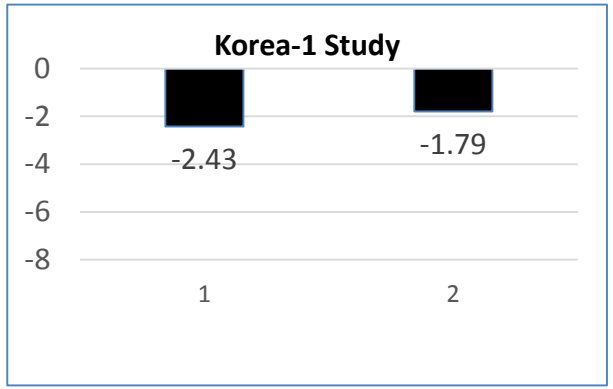
Study Sample Sizes

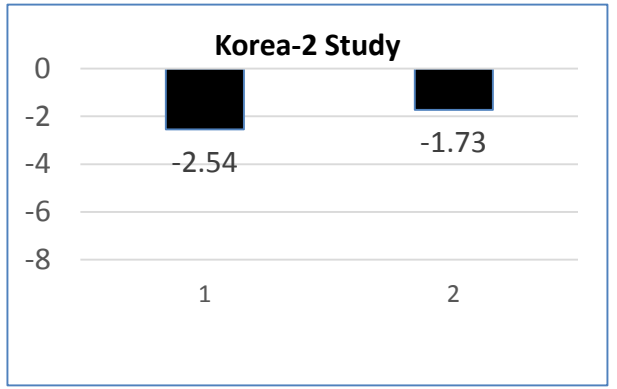


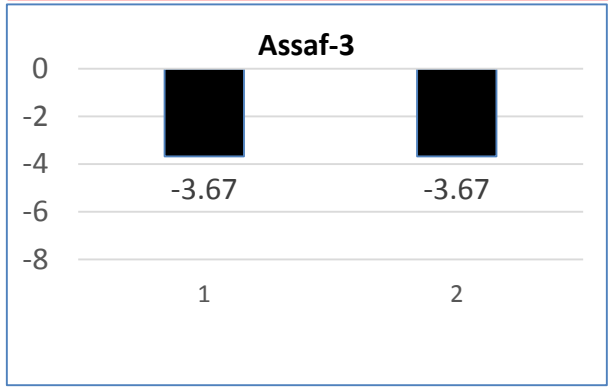
	1	2	3	4	5	6	7	8	9	10	11	12	13
Series1	60	18	11	8	7	10	10	6	10	10	4	6	84
Series2	50	8	11		8	6	6	2					
Series3							5	5					

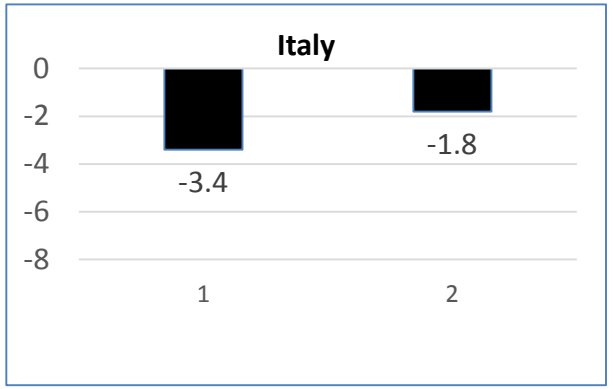
Mean Difference Between Groups Over Time, Baseline ADAS-Cog ≤30

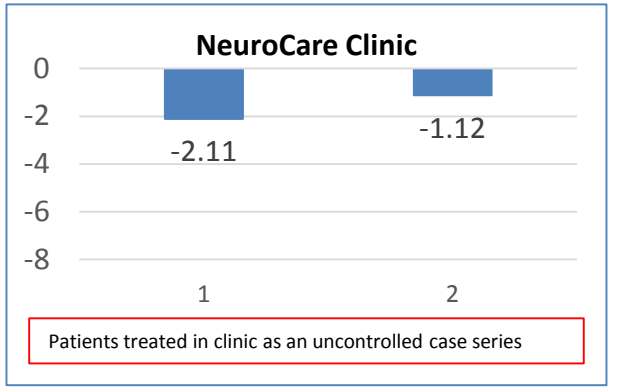










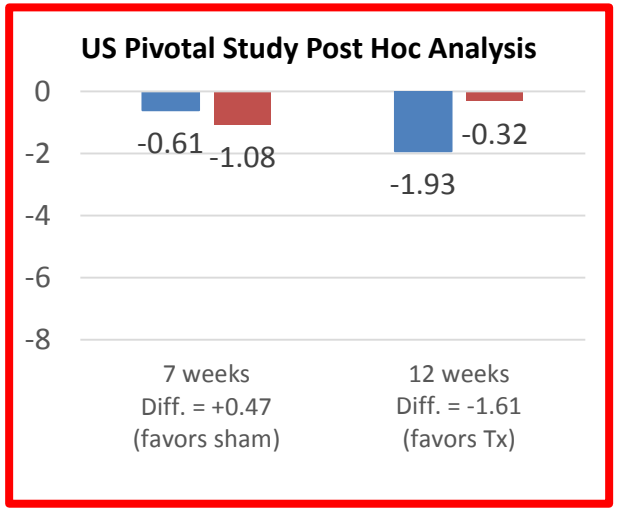


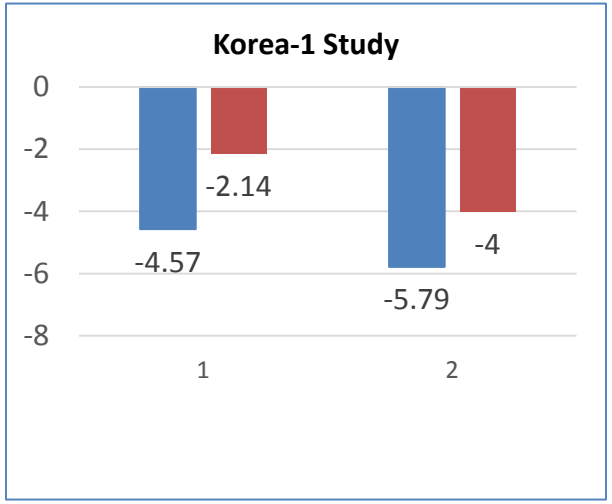
Only in the pivotal study does the magnitude of the between-group difference increase over time

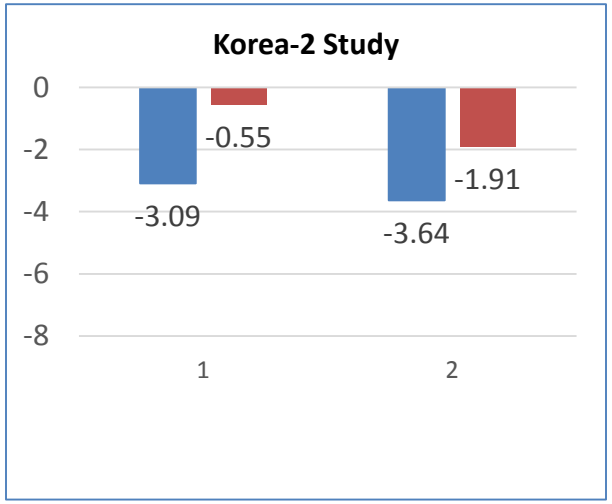
Note: These studies were selected because they had raw data at both follow-up timepoints

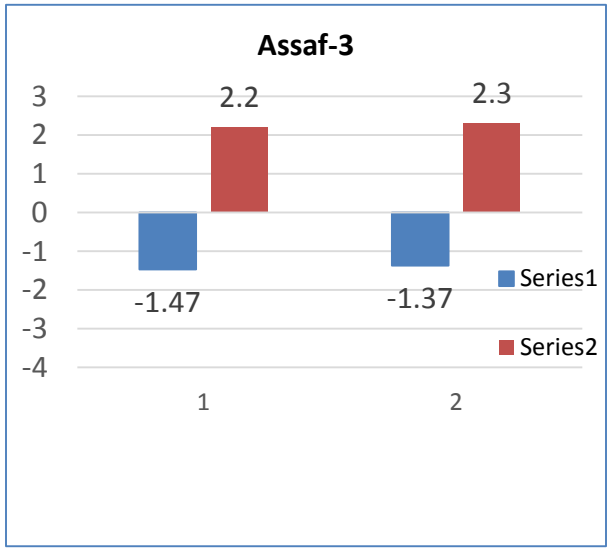
Sample Sizes	US Pivotal	Korea Pilot	Korea Pivotal	Assaf-3	Italy	NeuroCare
6-10 weeks	Active = 53 Sham = 48	Active = 14 Sham = 7	Active = 11 Sham = 11	Active = 9 Sham = 4	Active = 5 Sham = 1	Active = 32
10-14 weeks	Active = 51 Sham = 47	Active = 14 Sham = 7	Active = 11 Sham = 11	Active = 9 Sham = 4	Active = 5 Sham = 1	Active = 5

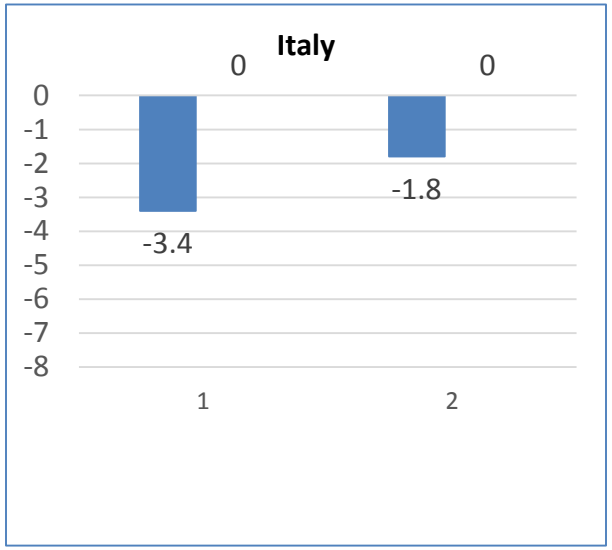
Mean Difference Between Groups Over Time, Baseline ADAS-Cog ≤30 (continued)

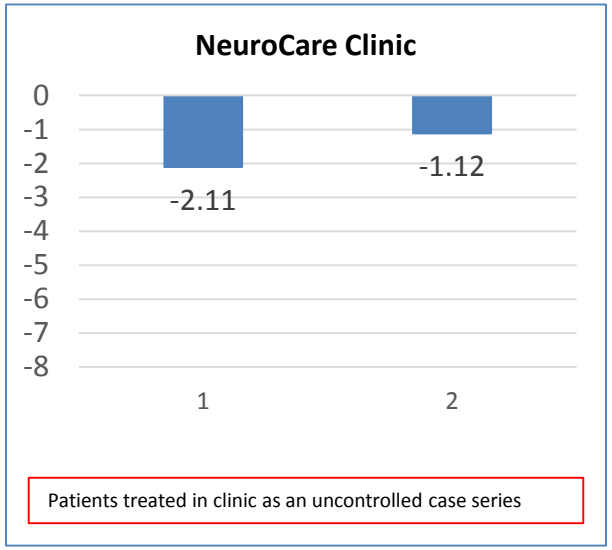












While the magnitude of the between group differences decrease over time in all but the pivotal study, the underlying data demonstrate inconsistent trends across studies.

Note: These studies were selected because they had raw data at both follow-up timepoints

Summary of Clinical Evidence

- High Uncertainty:
  - US pivotal pre-specified primary endpoint favored sham
  - US pivotal post-hoc analysis is hypothesis-generating
    - Baseline ADAS-Cog ≤30 subgroup outperforming entire cohort not verified in independent datasets
    - Trend from 7 weeks to 12 weeks not verified in independent datasets
- Best case result from US Pivotal Study in favor of treatment:
  - Baseline ADAS-Cog ≤30 subgroup
  - Mean difference between active and sham -1.61 at 12 weeks
- AEs are consistent with moderate-risk profile

Neuronix Ltd.'s neuroAD Therapy System: Statistical Considerations

Neurological Devices Advisory Panel Meeting
March 21st, 2019

Laura Thompson, PhD – Mathematical Statistician
Division of Biostatistics (DBS)
Office of Surveillance and Biometrics (OSB)
Center for Devices and Radiological Health (CDRH)
Food and Drug Administration

Outline: Statistical Considerations

- Review of study design and primary endpoint from a statistical perspective
  - ADAS-Cog score over assessment time
  - Interaction test between baseline ADAS-Cog and treatment group on change from baseline on ADAS-Cog
- Uncertainty of post-hoc subgroup indication
  - Baseline ADAS-Cog score of 30 was proposed after results were known
  - Cut-point comes from a measurement instrument
- Comments on Sponsor's meta-analysis

Review of Study Design Aspects

- Primary Endpoint: Change in ADAS-Cog score at 7 weeks
  - 12-week assessment recommended by FDA for durability of treatment effect only.

- Test for interaction between baseline ADAS-Cog and treatment group was prespecified
  - If not significant: any treatment effect would not be modified by baseline ADAS-Cog score.
  - If significant: No plan specified

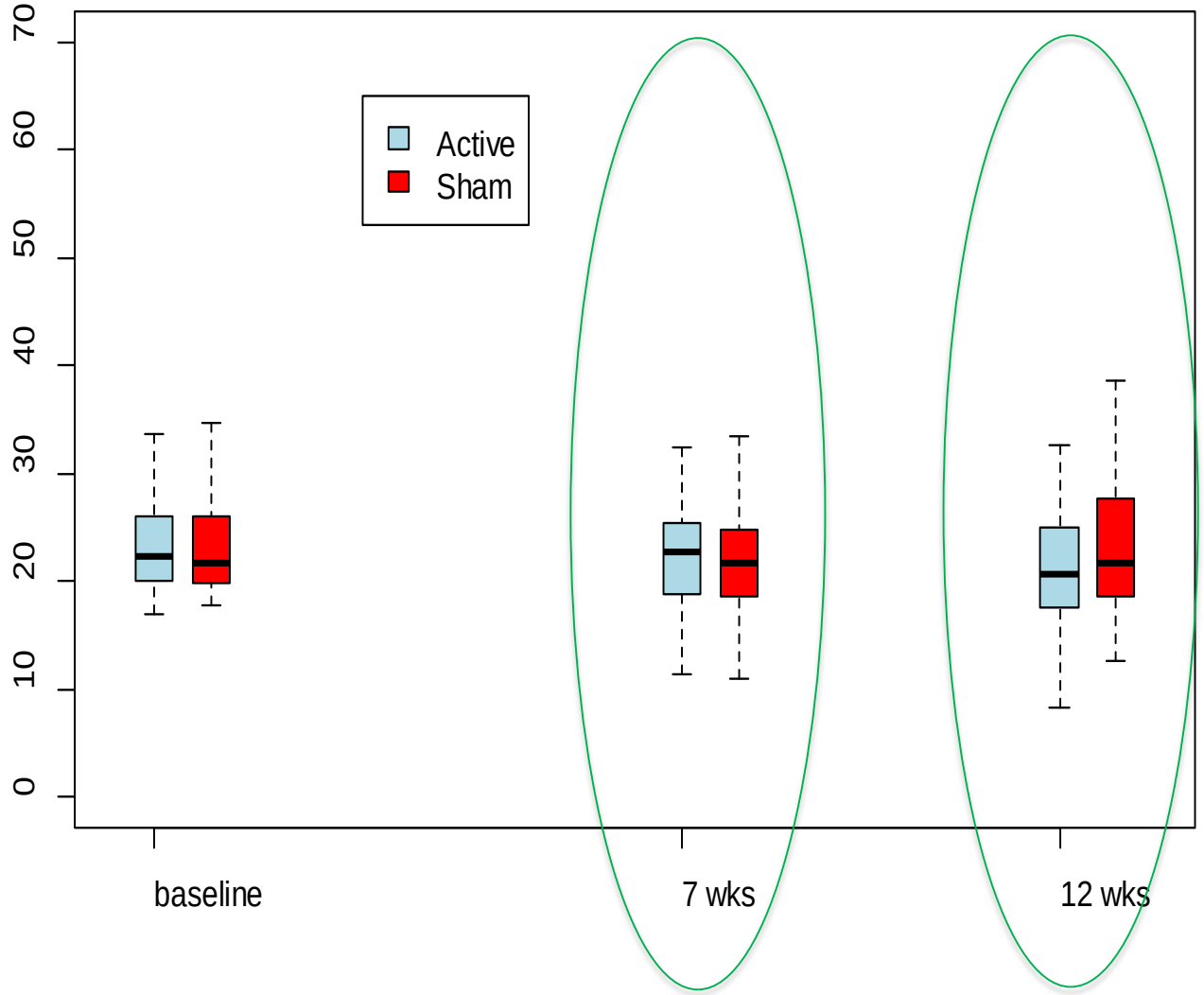
Primary Endpoint: Change in ADAS-Cog at 7 weeks
Secondary Endpoint: Change at 12 weeks

	Baseline ADAS-Cog	Change in ADAS-Cog at 7 weeks	Change in ADAS-Cog at 12 weeks
Active (n=53)	23.6 (SD = 4.8)	0.07	-1.03
Sham (n=48)	24.4 (SD = 6.5)	-1.38	-0.61
Difference (neg favors Active)		**+1.45** (95% CI: -0.27, 3.16)	**-0.42** (95% CI: -2.19, 1.35)
p-value	0.93	0.09	0.64

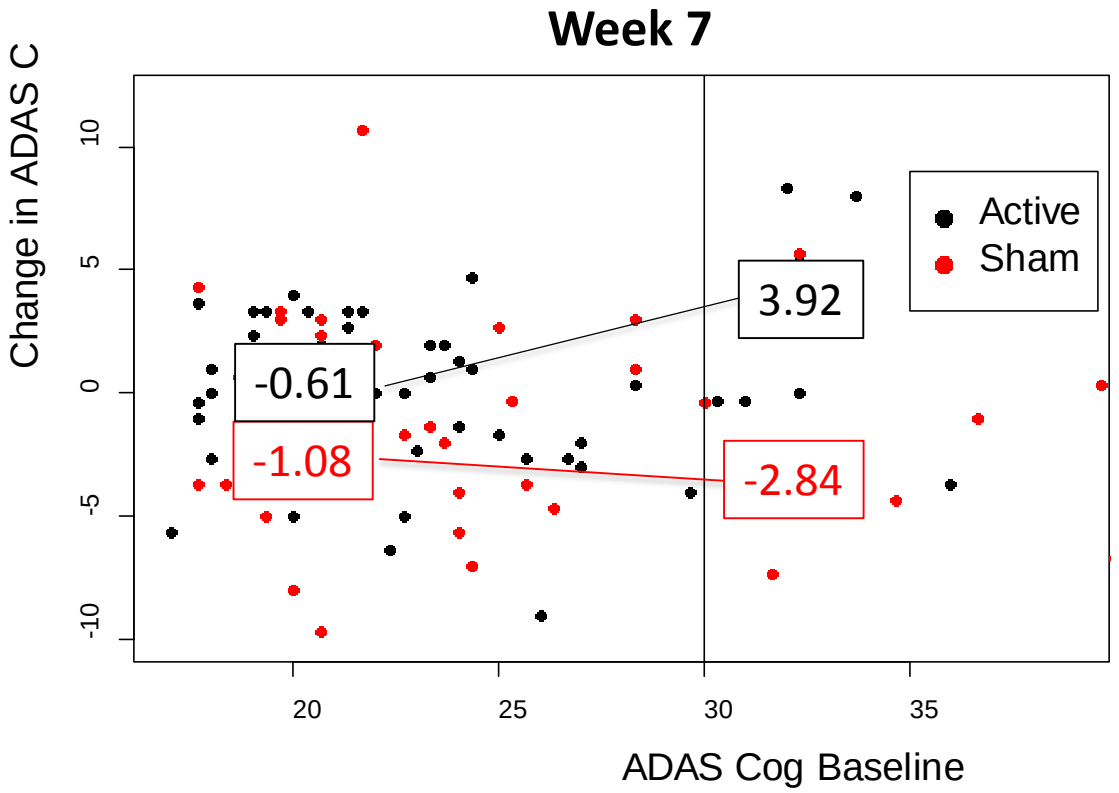
Pivotal Study: ADAS-Cog Score by Assessment Time (All Subjects)

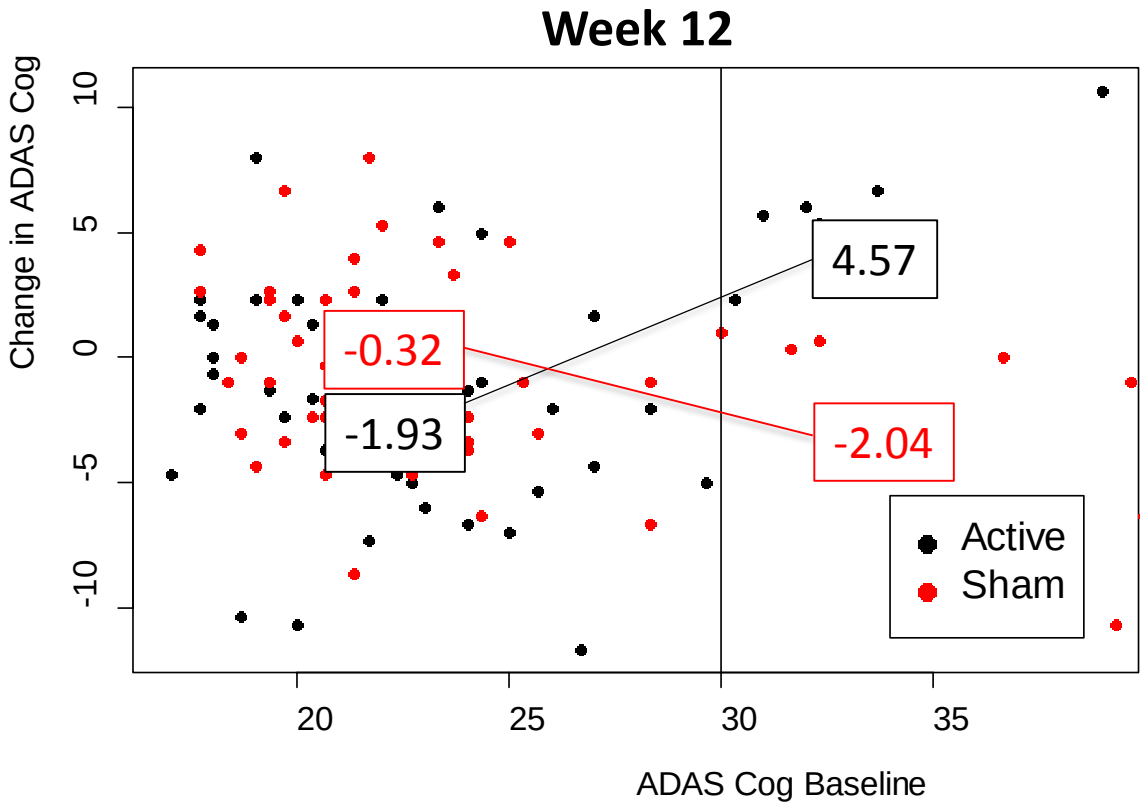
Worse Score

Better Score



Interaction between Treatment Group and Baseline ADAS-Cog on Change in ADAS-Cog



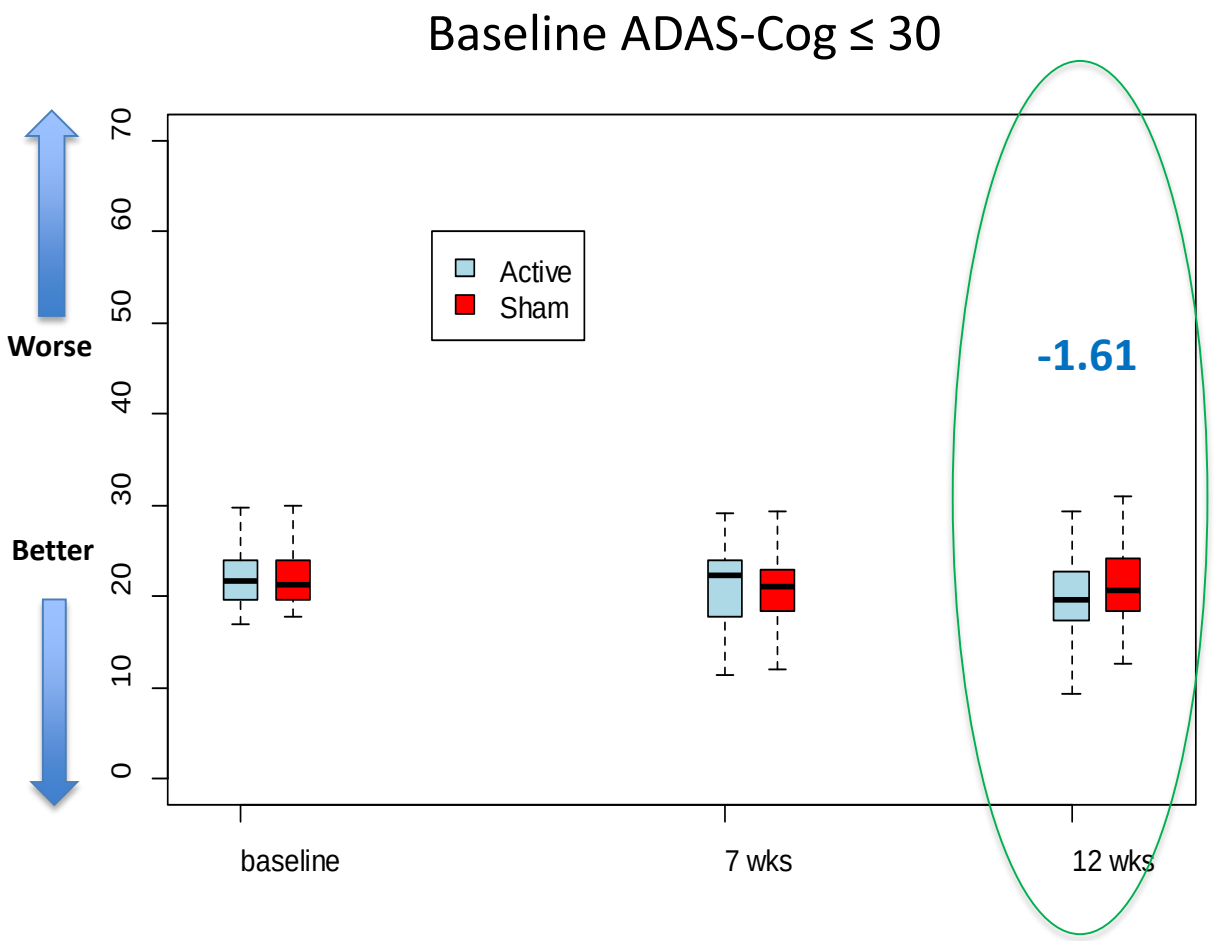


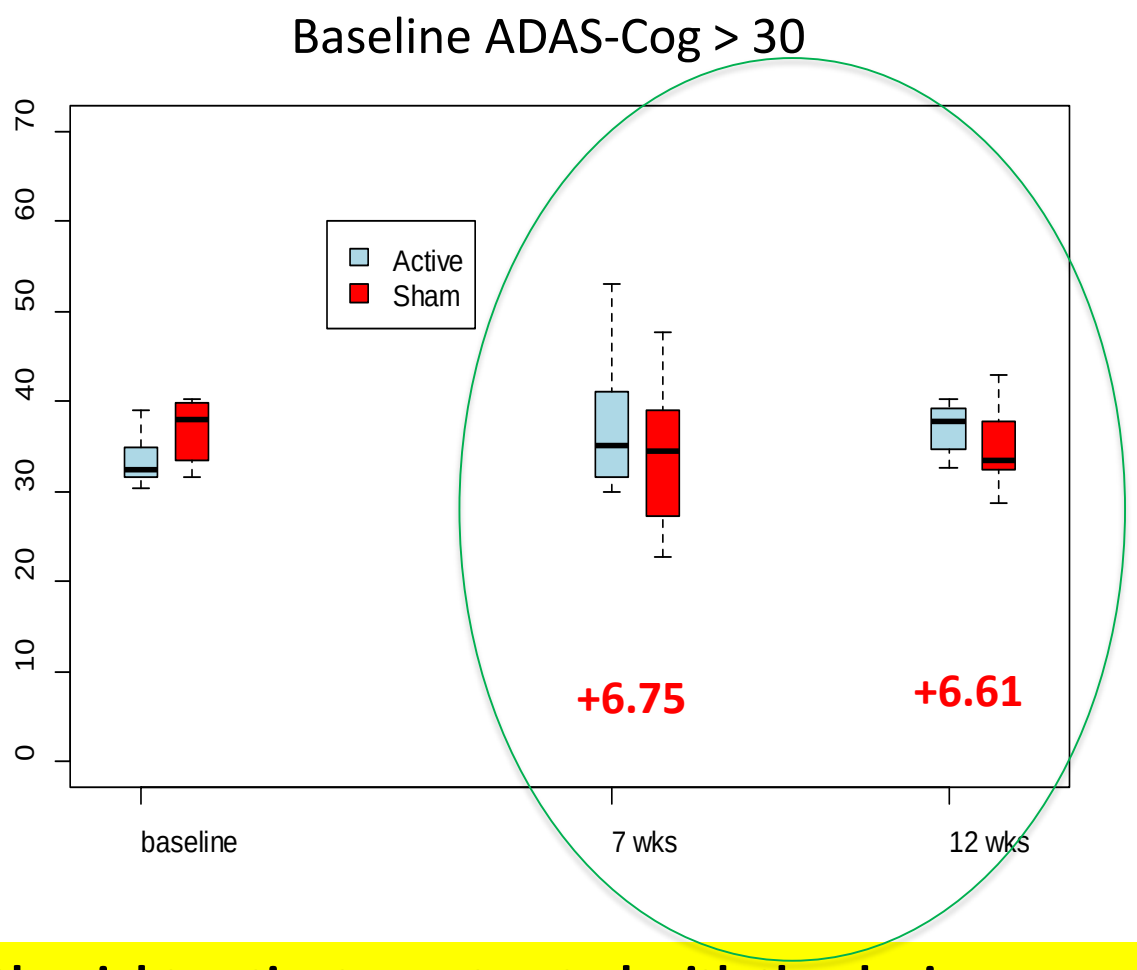
Evidence of Qualitative interaction - Active is effective in one subgroup, but the sham is effective (over Active) in the other subgroup. Qualitative IA is relatively rare.

Pivotal Study: Change in ADAS-Cog by Subgroup

Subjects with baseline ADAS-Cog <= 30		
	Change in ADAS-Cog (7 weeks)	**Change in ADAS-Cog (12 weeks)**
Active (n=45)	-0.61 (SD=3.13)	-1.93 (SD=4.27)
Sham (n=40)	-1.08 (SD=3.81)	-0.32 (SD=3.87)
Active – Sham	**+0.47 (p = 0.53)** 95% CI: (-1.02, 1.96)	**-1.61 (p = 0.08)** 95% CI: (-3.35, 0.13)
Subjects with baseline ADAS-Cog > 30		
	Change in ADAS-Cog (7 weeks)	**Change in ADAS-Cog (12 weeks)**
Active (n=8)	3.92 (SD=5.96)	4.57 (SD=4.76)
Sham (n=8)	-2.84 (SD=7.72)	-2.04 (SD=4.36)
Active – Sham	**+ 6.75 (p = 0.07)** 95% CI: (-0.69, 14.20)	**+6.61 (p = 0.02)** 95% CI: (1.47, 11.76)

Pivotal Study: ADAS-Cog Score Over Time





The certainty of the cut-point may be important to ensure the right patients are treated with the device.

Post-hoc Subgroup Selection (Baseline ADAS-Cog ≤ 30) and Data-driven Hypothesis Tests

- A pre-specified interaction involving a continuous covariate *does not pre-specify* a cut point for a subgroup.
- Analyses associated with a post-hoc subgroup carry greater uncertainty than prospective hypothesis tests.
  - The nominal p-value may be incorrect
  - Chance of finding a significant p-value increases with number of tests done
  - Biases in choosing a hypothesis that gives a good test result.
  - Likelihood of a good result by chance is higher when a result is highlighted post-hoc
- Several different post-hoc subgroups might be identified that could separate the population into responders and non-responders, even using independently generated covariates.

Uncertainty Regarding "30" as the Cut-Point

- It is not a natural subgroup such as that determined by age or gender.
  - It was derived using a measurement; the cut-point used to indicate the population for the device contains measurement error.
- A *range* of values around 30 may be clinically equivalent to a score of 30, especially across different sites or days.
- Tolerance for uncertainty may be lower in order to be sure that 30 is an appropriate value that will not prevent treatment for those who could have benefitted or recommend treatment for those who may worsen

FDA Analysis of Baseline ADAS-Cog ≤30 as the Cut-Point

- FDA investigated the validation of the baseline ADAS-Cog value of 30.
- Can this cut point distinguish responders to neuroAD over sham from non-responders, in an independent data set (preferably external to the study)?
- FDA estimated the treatment effect for baseline ADAS-Cog ≤30 using the supplemental studies, and compared that result to the treatment difference in the full study cohort in order to evaluate an enhanced effect in the subgroup.
- Key question: Was the enhanced effect in the US subgroup merely a chance finding from the pivotal study?

External Test of the Baseline ADAS-Cog ≤30 Cut-Point

	Korea 1 Subgroup (14 Active; 7 Sham)	Italy Subgroup (5 Active; 1 Sham)	Korea 1 + Italy (19 Active; 8 Sham)	Korea 2 Study (11 Active; 11 Sham)
Treatment diff in subgroup	–1.79 (–5.39, 1.81)	–1.80 (NA)	–1.24 (–4.50, 2.02)	-1.73 (-4.74, 1.28)
Treatment diff in entire study	-2.51	+0.17	-2.03	-1.73
Observed enhanced difference of subgroup over the entire population	–1.79 – (–2.51) = +0.72	–1.80 – (+0.17) = – 1.97	-1.24– (-2.03) = +0.79	NA

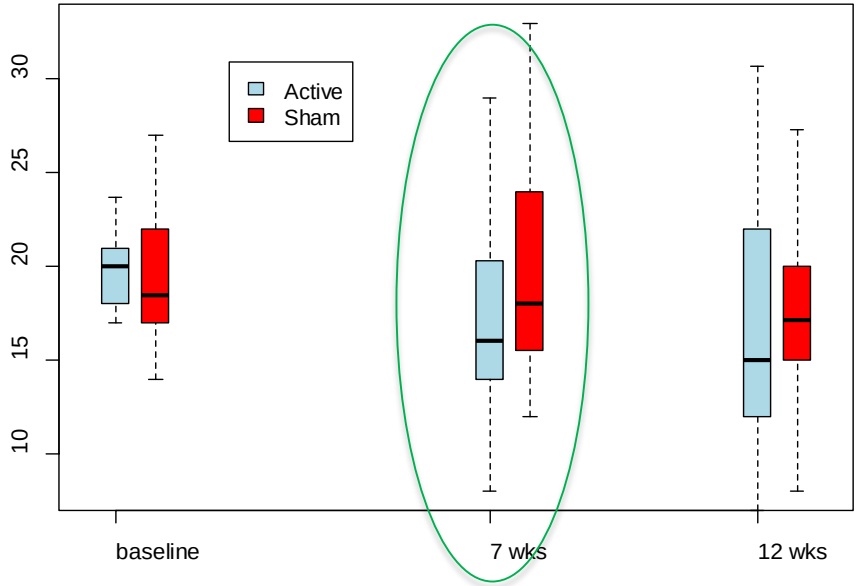
A *decreased* result over the entire population at 12 weeks was also seen in the pooled supplemental studies

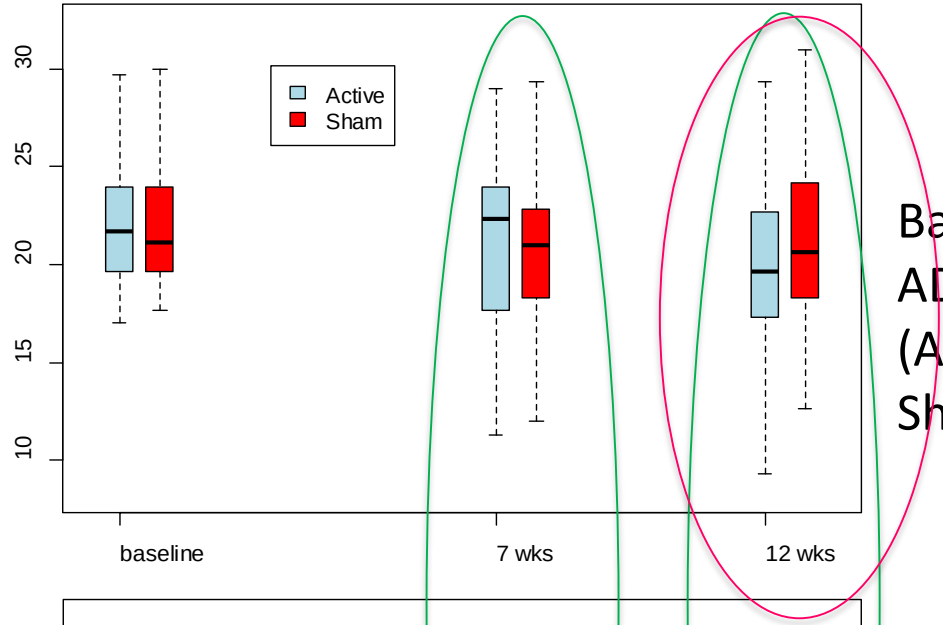
General Inconsistency in ADAS-Cog Over Time and Subgroup

Supplemental RCTs Combined – Korea, Assaf3, Italy

US Study

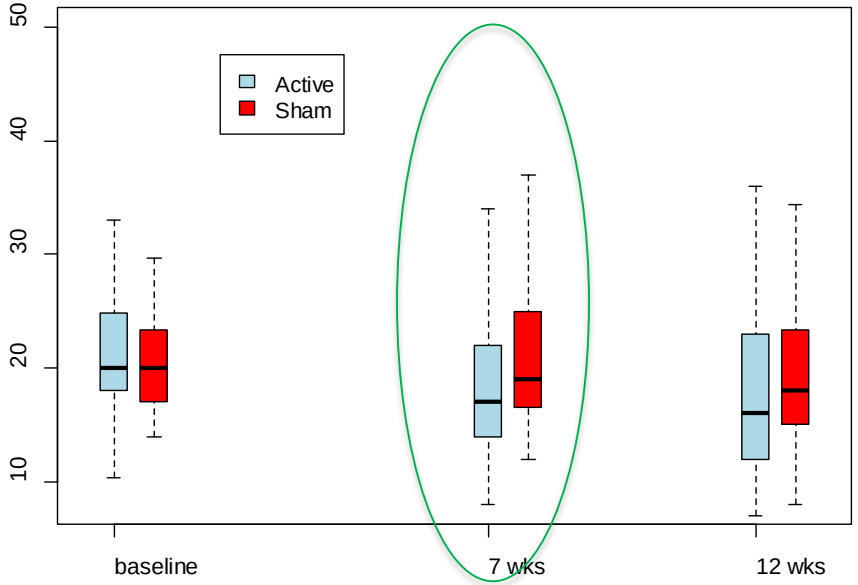
Baseline ADAS-Cog<30 (Active n=38 Sham n=24)

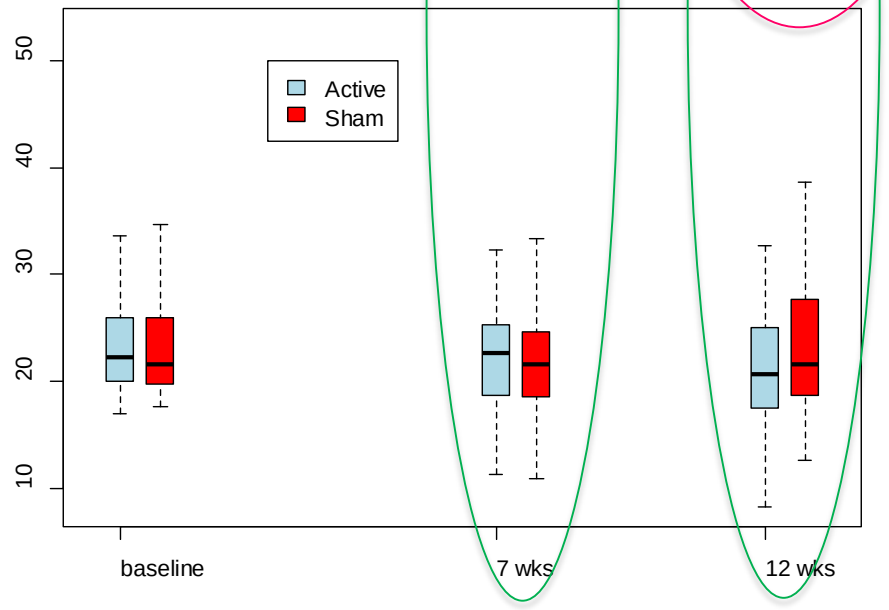




Baseline ADAS-Cog<30 (Active n=45 Sham n=40)

Full Cohort (Active n=43 Sham n=27)





Full Cohort (Active n=53 Sham n=48)

External Test of the Baseline ADAS-Cog ≤30 Cut-Point

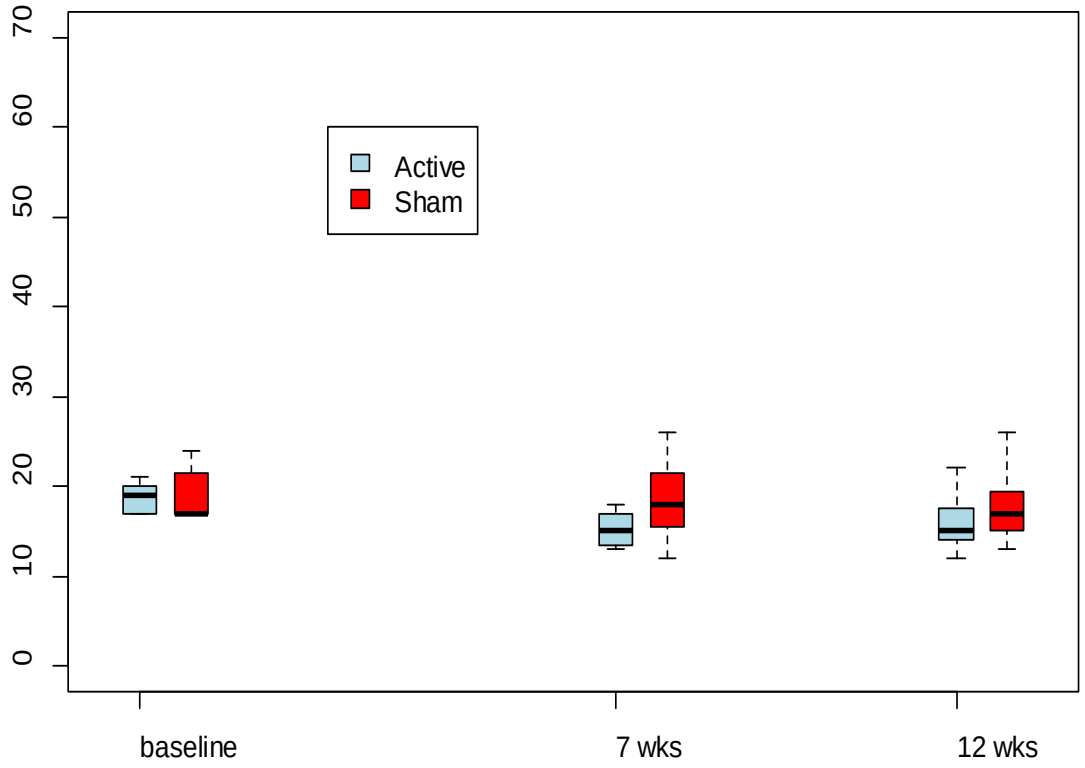
	Korea 1 Subgroup (14 Active; 7 Sham)	Italy Subgroup (5 Active; 1 Sham)	Korea 1 + Italy (19 Active; 8 Sham)	Korea 2 Study (11 Active; 11 Sham)
Treatment diff In subgroup	–1.79 (–5.39, 1.81)	–1.80 (NA)	–1.24 (–4.50, 2.02)	-1.73 (-4.74, 1.28)
Treatment diff in entire study	-2.51	+0.17	-2.03	-1.73
Observed enhanced difference of subgroup over the entire population	–1.79 – (–2.51) = +0.72	–1.80 – (+0.17) = – 1.97	-1.24– (-2.03) = +0.79	NA

Concerns With Considering Korea-2 as "Confirmatory"

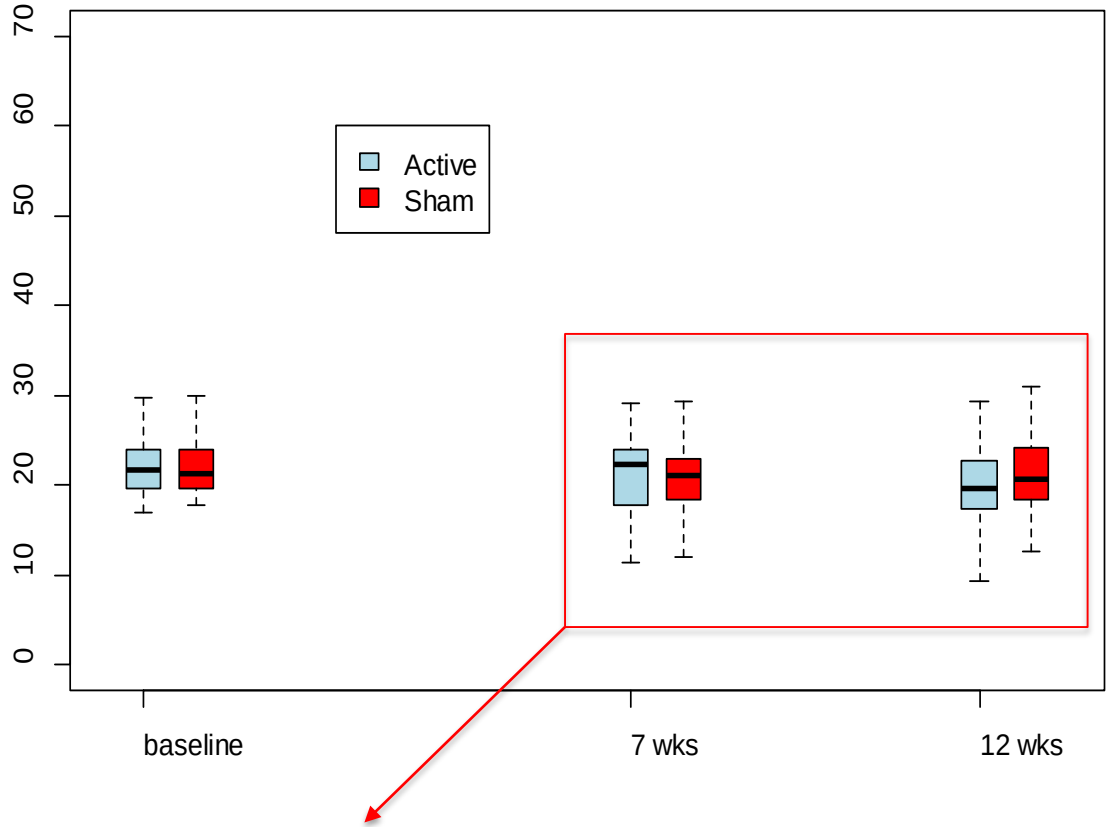
- Enrolled only 22 subjects so far. Treatment difference has substantial variability: 95% CI: (-4.74, 1.28)
- Pattern of results across assessment times different from US study
  - An increase in treatment difference from 7 to 12 weeks despite no treatment applied after 7 weeks was not seen in Korea-2 (nor often in the supplemental studies).
- Korea-2 study may not generalize to US

Korea-2 vs. US subgroup
ADAS-Cog Score By Assessment Time

Korea-2 (baseline ADAS-Cog<=30)



US subgroup baseline ADAS-Cog<=30



FDA has not been able to determine why the treatment difference would increase after ceasing stimulation

Sponsor's Meta-Analysis

- Studies used: US study, Korea 1, Korea 2
- Only subjects with baseline ADAS-Cog <= 30 were used.
- Exchangeability among the 3 studies is assumed.
- FDA does not agree that the Korea studies are exchangeable with the US study:
  - Results from Korea studies show different pattern of effectiveness of neuroAD from 7 weeks to 12 weeks.
  - Different countries
  - Evidence of different average motor thresholds which may impact response to TMS.

Confirming a Post-Hoc Subgroup Result: CDRH Practice

- Sponsors should collect additional data to “confirm” subgroup effect.
- A new analysis might statistically combine the results from the two sources, but should not eliminate the poorer performing subgroup(s) from the original study
- Poorer-performing subgroup(s) may serve as a multiplicity adjustment to down-weight potentially spurious results from the post-hoc finding.

Statistical Conclusions

- Pre-specified hypotheses that are documented and planned to be tested before any examination of the data are a tenet of good trial design.
- Despite the interaction test being pre-specified, the sponsor's intention was to make a claim for the effectiveness of neuroAD over sham for the entire population.
- The cut-point was chosen after study data were available, and then tested on those data. Hence the subgroup result may be inflated; p-value not adjusted for multiplicity
- The supplemental studies collectively showed different patterns of treatment benefit over sham than did the US study across the two follow-up assessments, as well as across the <30 and >30 subgroups.
- Taken together, these points lead to uncertainty in concluding device effectiveness in a US population.

Neuronix Ltd.'s neuroAD Therapy System
Benefit-Risk Discussion

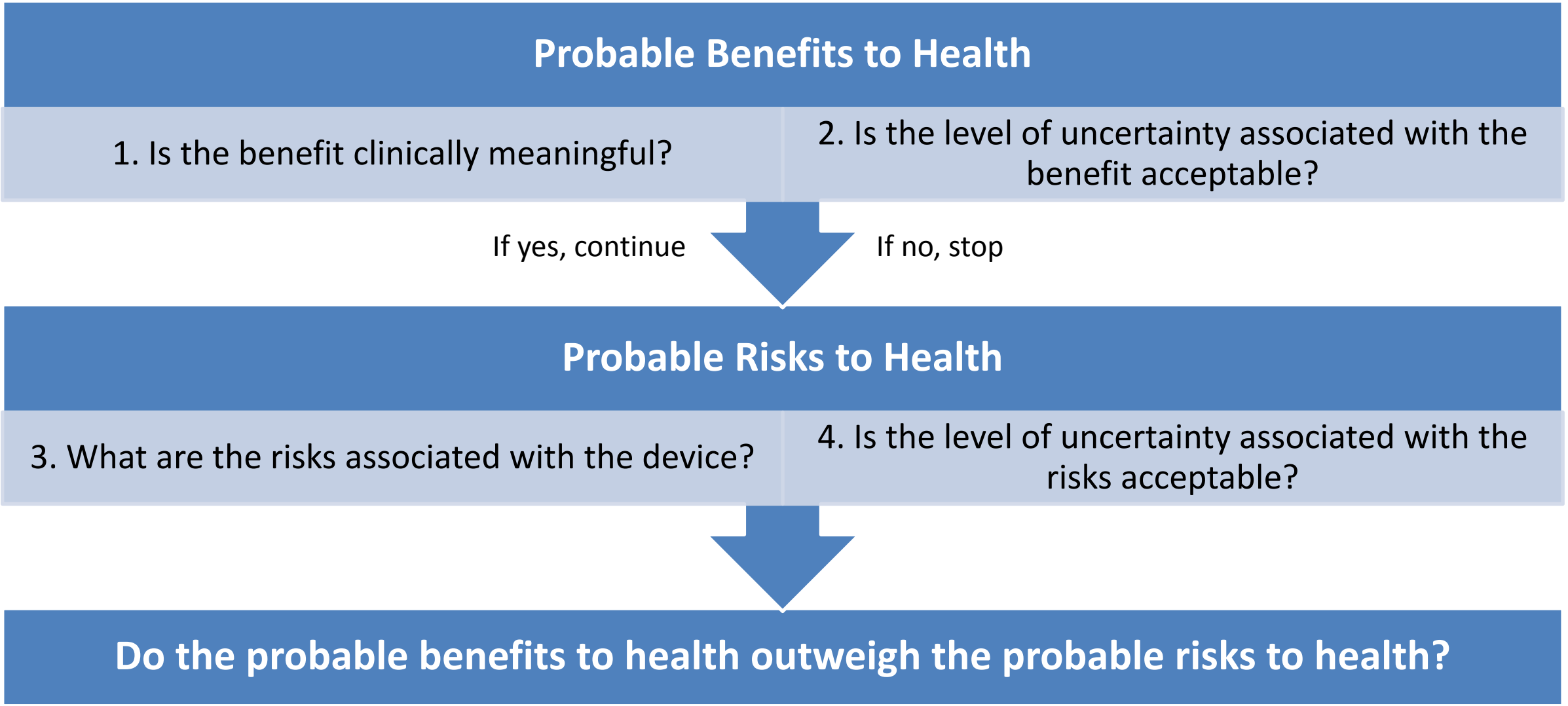
Neurological Devices Advisory Panel Meeting
March 21st, 2019

Claudette Brooks, MD – Neurologist and Clinical Reviewer
Division of Neurological and Physical Medicine Devices (DNPMD)
Office of Device Evaluation (ODE)
Center for Devices and Radiological Health (CDRH)
Food and Drug Administration

Outline: Benefit-Risk Discussion

- Characterizing Clinically Meaningful Benefit
- Characterizing Risks
- Benefit-Risk Summary

Benefit-Risk Assessment Process, Simplified



Review of Device Effectiveness Results

- Sham outperformed active treatment by 1.45 points at the primary endpoint in the pivotal study (7 weeks).

- The largest observed difference in ADAS-Cog that favors the device was 1.61 points in the baseline ADAS-Cog ≤30 subgroup at 12 weeks

- The largest observed difference in CGI-C favors the active treatment but does not exceed 0.5 points.
  - Looking at all possible data points, i.e., 7 and 12 weeks, overall cohort or post-hoc subgroup baseline ADAS-Cog<=30

Minimum Clinically Important Difference (MCID)

- No consensus in the Alzheimer's Disease clinical community
- Literature: Range of MCID opinions, 2-5 points
- Network of Experts input: 4-5 points for symptomatic treatment, 2 points for disease modifying therapy
- Sponsor's Physician Survey: > 50% of physicians considered at least 2 points or greater on the ADAS-Cog score to be clinically meaningful following 3 months of treatment. This was also the largest consensus.

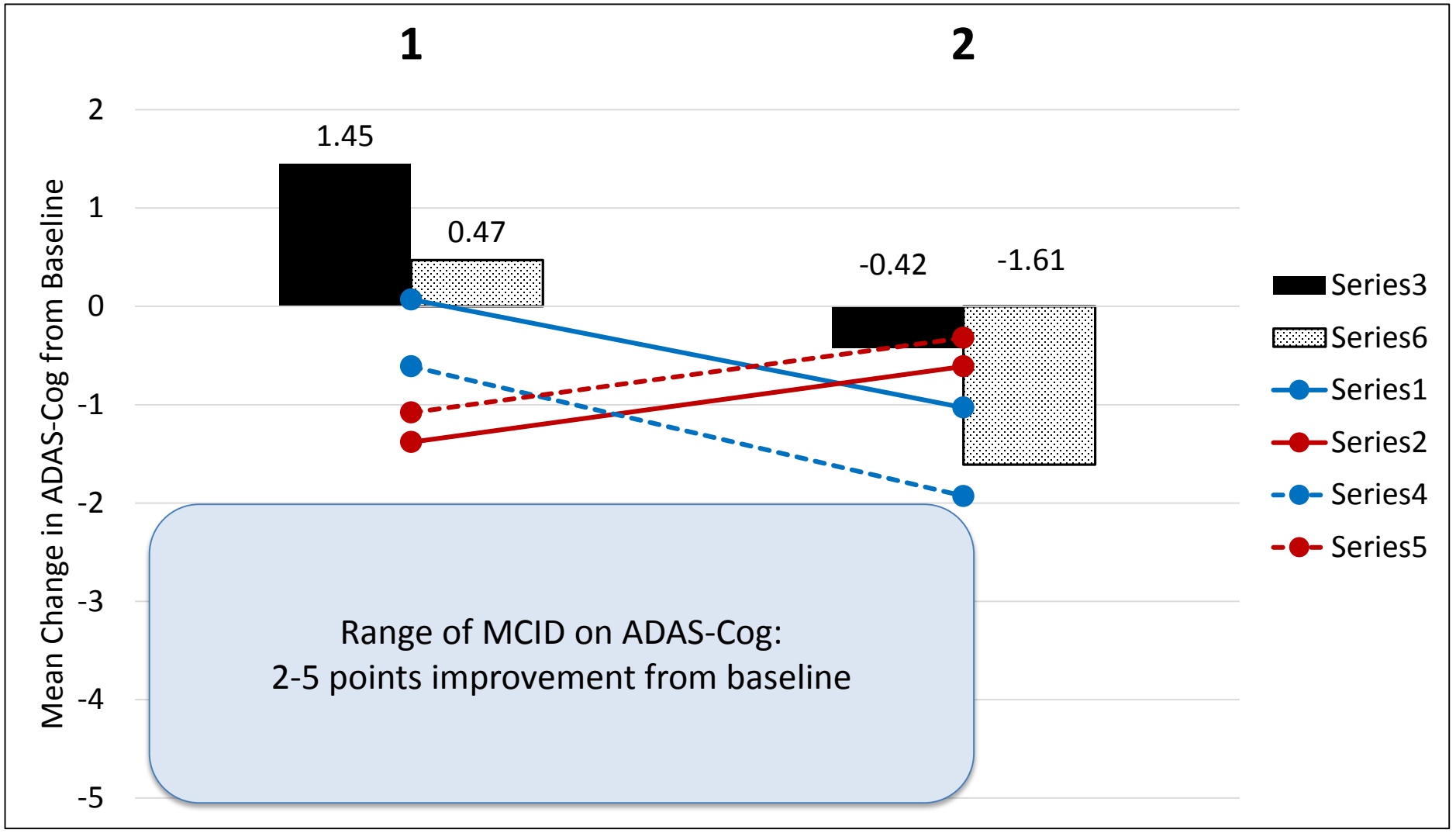
Example: What does a 2 point change mean?

ADAS-Cog ex: Testing Word Recall

- 3 trials to learn 10 words
- Score= mean # of words NOT recalled on each trial (max score 10)
- 2 additional words from previous trial → 2 point improvement on the overall ADAS-Cog score
- Unclear how this translates to global function

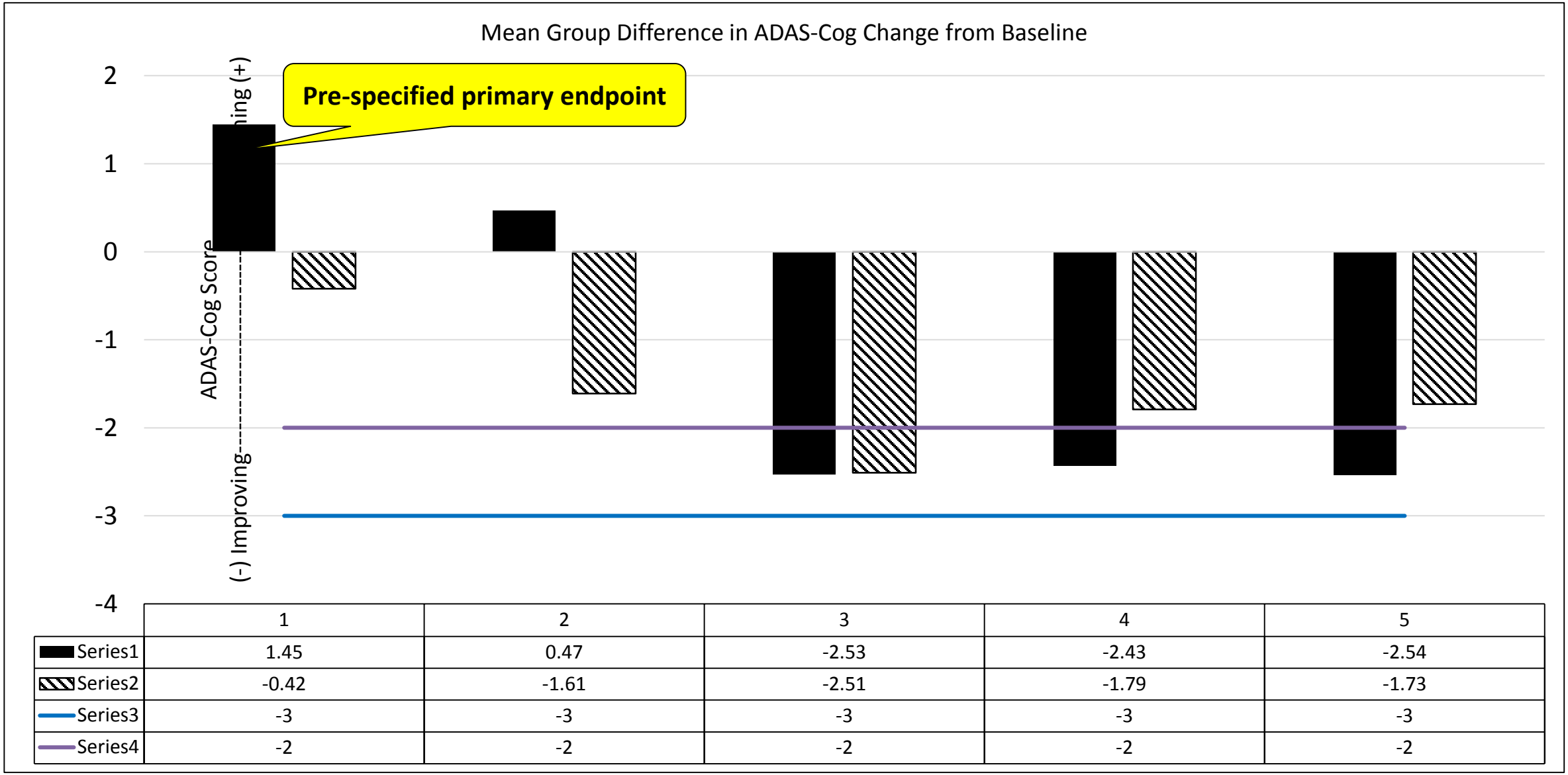
May be influenced by external variables or confounders, e.g., serial testing/practice effects, rater experience, patient mood (psychometrics)

MCID in Context – Pivotal Study



The largest difference between groups is 1.61 points and occurs at 12 weeks in the subgroup

MCID In Context – All Data Sources



	1	2	3	4	5
Series1	1.45	0.47	-2.53	-2.43	-2.54
Series2	-0.42	-1.61	-2.51	-1.79	-1.73
Series3	-3	-3	-3	-3	-3
Series4	-2	-2	-2	-2	-2

CAUTION: Differences with Approved Drugs

Approved Treatments	neuroAD
• Cognitive and functional (or global) co-primary endpoints • Statistically significant difference between drug and placebo on both endpoints (approvals do not rely on a MCID) • Results replicated in at least two adequate and well controlled clinical trials with hundreds of subjects	• ADAS-Cog was sole primary endpoint • Sham outperformed active treatment at pivotal study primary endpoint. • One pivotal study and a collection of smaller studies with sample sizes <30

The process for drug approvals and the results on which those approvals are based are different, and any comparison needs to account for both.

Other Variables, Reliability, and Measurement Error

- Cognitive Testing: Serial Testing and Practice Effects
- ADAS-Cog Test-Retest Reliability (0.9-0.93)
  - Depending on the SD could correspond to 1-3 points in measurement error*
- ADAS-Cog Measurement Error
  - Variability about 3 - 4 for a site that is not well-trained, and about 1.5 - 2 for a site that is well-trained

*Anzalee Khan, et al, "Reliability of the Alzheimer's Disease Assessment Scale (ADAS-Cog) in longitudinal Studies", Current Alzheimer Research (2013) 10: 952. Mohs, R, et al, "A new rating scale for Alzheimer's disease" Am J Psychiatry 141: 1356-1364. Robert L. Heilbronner Ph.D. , et al.(2010) Official position of the American Academy of Clinical Neuropsychology on serial neuropsychological assessments: the utility and challenges of repeat test administrations in clinical and forensic contexts, The Clinical Neuropsychologist, 24:8, 1267-1278,

Uncertainty in Totality of Clinical Evidence

- Pivotal Study Primary Endpoint:
  - Sham outperformed Active
- Pivotal Study Post-Hoc Subgroup: Hypothesis-generating
  - ADAS-Cog<=30 subgroup not verified in supplemental datasets
  - Delayed effect (12 weeks) not verified in supplemental datasets
  - Change of 1.61 not clinically meaningful; MCID of 2-5 points
- Concerns with using ADAS-Cog to select patients
  - Continuous cut-off, measurement error
  - ADAS-Cog>30 poor performers

CHARACTERIZING RISKS

General Risks neuroAD

- Seizure
- Thermal Injury
- Hearing Loss
- Scalp Discomfort, dizziness, nausea, pain in the neck or jaw, headache, or other AEs due to treatment
- Adverse Tissue Reaction
- Electrical Shock
- Device failure due to interference with other devices

General Risks Associated with TMS of other indications

- Unprovoked seizure
- Apraxia
- Sleep disturbances
- Olfactory dysfunction

Increased Risks Associated with Alzheimer's Disease Population vs General Population

Known Probable Risks Based on Adverse Event Data

- Pivotal Study:
  - headache, neck pain, skin discomfort or muscle twitching

- Supplemental Investigations:
  - psychiatric symptoms that required medication (n=1, deemed unrelated to the device per the PI), mild and transient hearing impairment post-intervention, blurry vision eye pain, neck pain/stiffness, mild scalp pain, soreness at stimulation site, achiness, fatigue, nausea, transient eye heaviness, mild to moderate headache events, tiredness, dizziness, increased anxiety

FDA Benefit-Risk Assessment

- Results of the clinical evidence do not appear to demonstrate a clinically meaningful benefit
  - We have low uncertainty in the primary effectiveness result which favored sham
  - We have significant uncertainty in the scientific validity of the results of post-hoc analyses (e.g., could be due to chance)

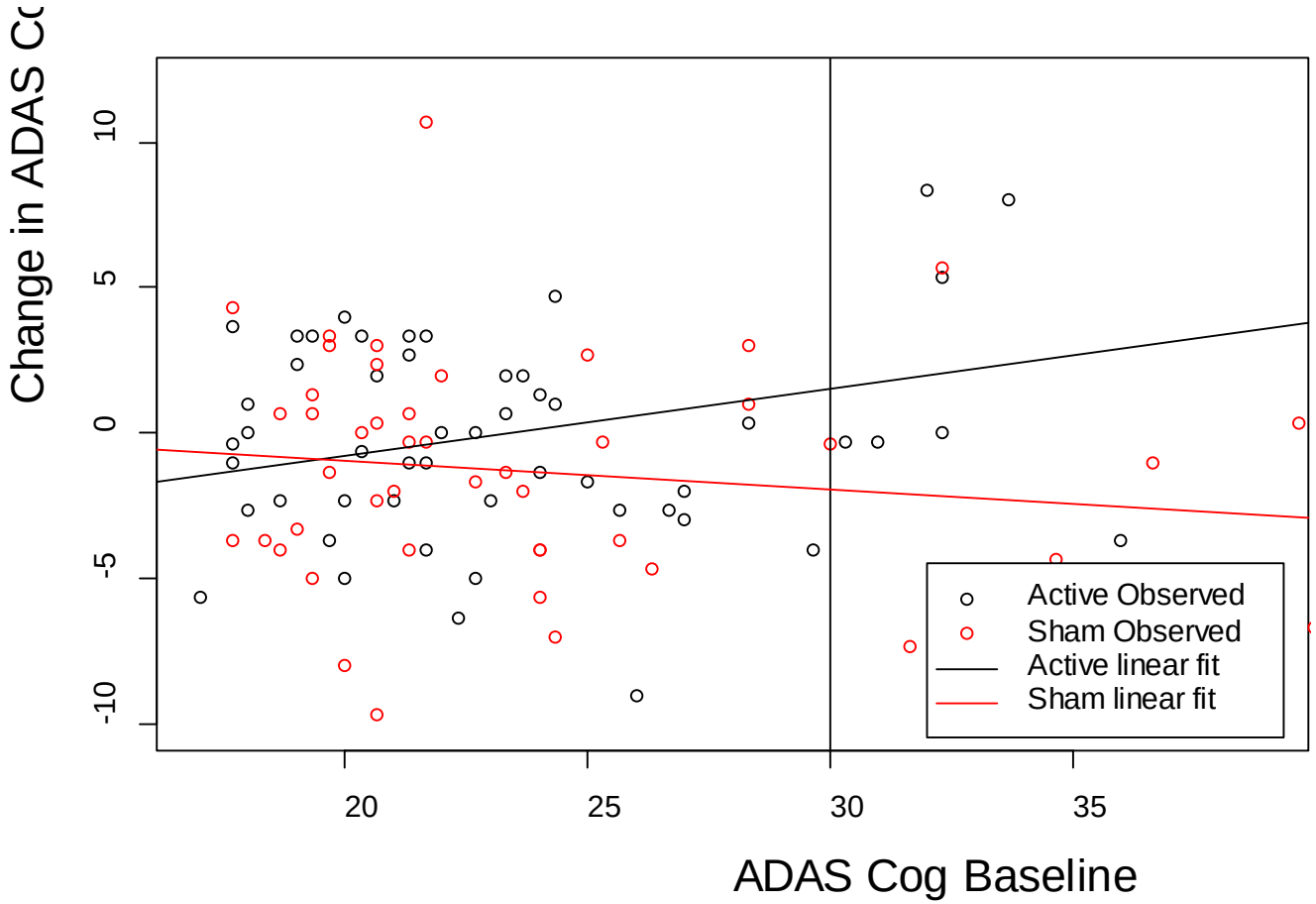
- While the risks appear to be moderate based on available information, we do not have the full safety data for the supplemental investigations

Closing Remarks

U.S. FOOD & DRUG ADMINISTRATION

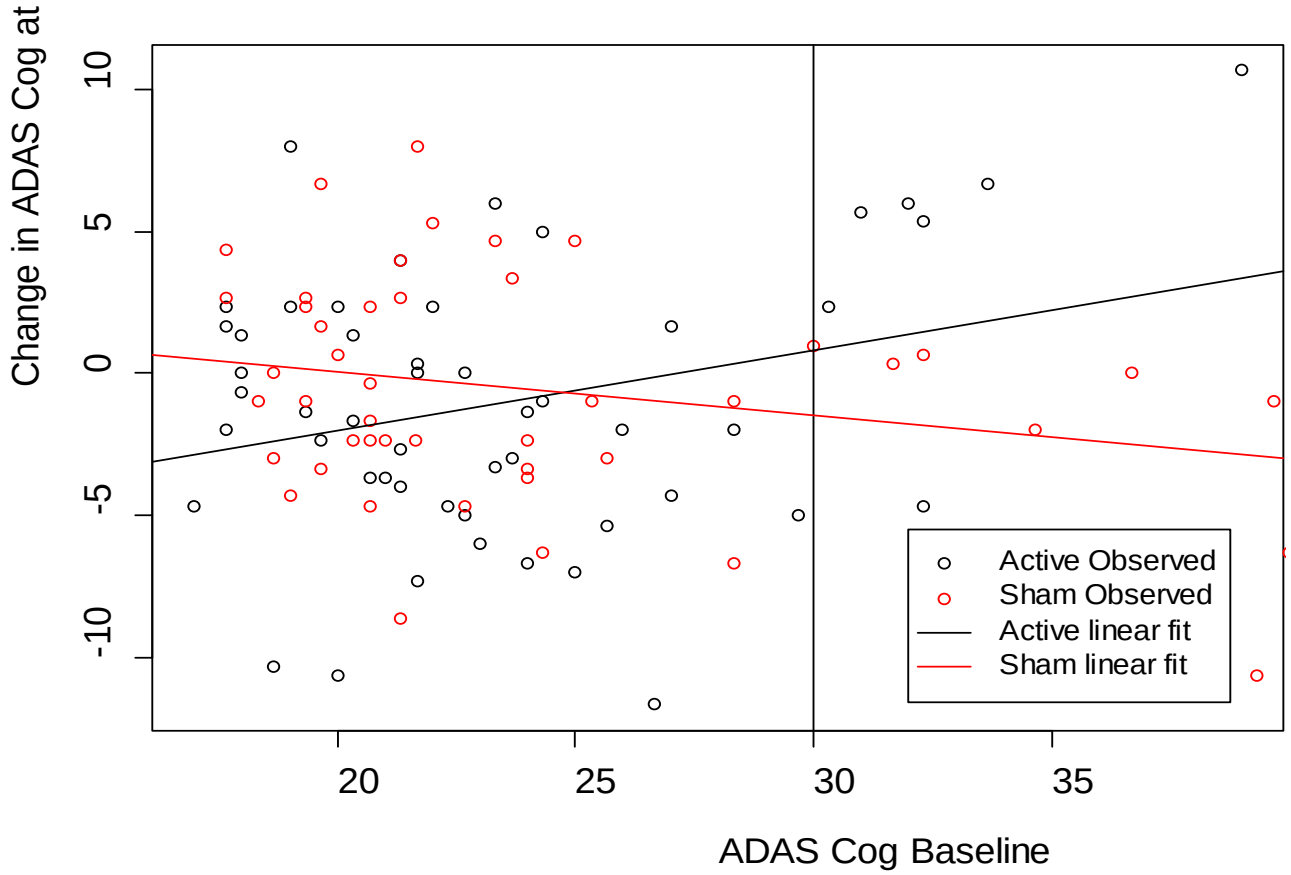
THANK YOU

Interaction between Treatment Group and Baseline ADAS-Cog on Change in ADAS-Cog at 7 weeks



Interaction p = 0.03

Interaction between Treatment Group and Baseline ADAS-Cog on Change in ADAS-Cog at 12 weeks



Interaction p = 0.007

Which treatment arm shows a “consistent” average trend, regardless of subgroup?

Subjects with baseline ADAS-Cog <= 30		
	Change in ADAS-Cog (7 weeks)	Change in ADAS-Cog (12 weeks)
Group 1 (n=45)	-0.61	-1.93
Group 2 (n=40)	-1.08	-0.32
Group 1 – Group 2	+0.47 (p = 0.53)	-1.61 (p = 0.08)
Subjects with baseline ADAS-Cog > 30		
	Change in ADAS-Cog (7 weeks)	Change in ADAS-Cog (12 weeks)
Group 1 (n=8)	3.92	4.57
Group 2 (n=8)	-2.84	-2.04
Group 1 – Group 2	+ 6.75 (p = 0.07)	+6.61 (p = 0.02)

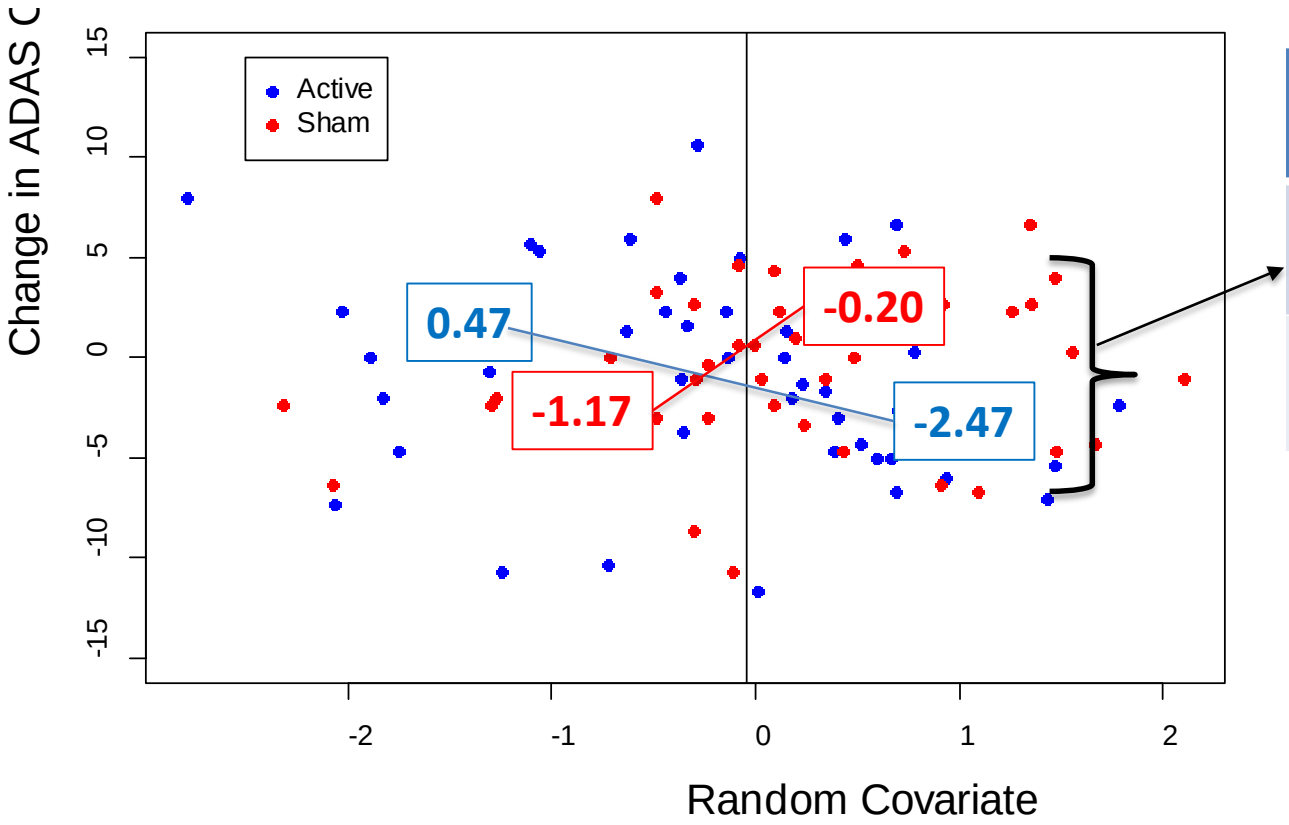
Two Types of Interaction (IA)

1. **Quantitative IA** - Active may be more effective in one subgroup over another, but both subgroups show the same trend (e.g., superiority of active over sham)
   - A marketing label might be modified to reflect a stronger benefit in one subgroup.

2. **Qualitative IA** - Active is effective in one subgroup, but the sham is effective (over Active) in the other subgroup
   - If the overall effect of the device is close to zero (overall difference close to 0), then a qualitative IA may be meaningless:
     – One can always divide the subjects by a subgroup into + and - treatment effects by finding a cut-point that separates non-responders from responders to the Active over control.

We have IA type #2: a qualitative IA with small/negligible overall difference (-0.42), with a “positive” subgroup (baseline ADAS-Cog < 30) and a “negative” (> 30) subgroup.

A "Significant Finding" From a Post-Hoc Analysis May Be Due to Chance: Random Covariate



Subgroup	Treatment Diff at 12 wks (95% CI)
Random Covariate > -0.043	**-2.27 (-4.40, -0.16)**
Random Covariate < -0.043	1.64 (-1.26, 4.52)

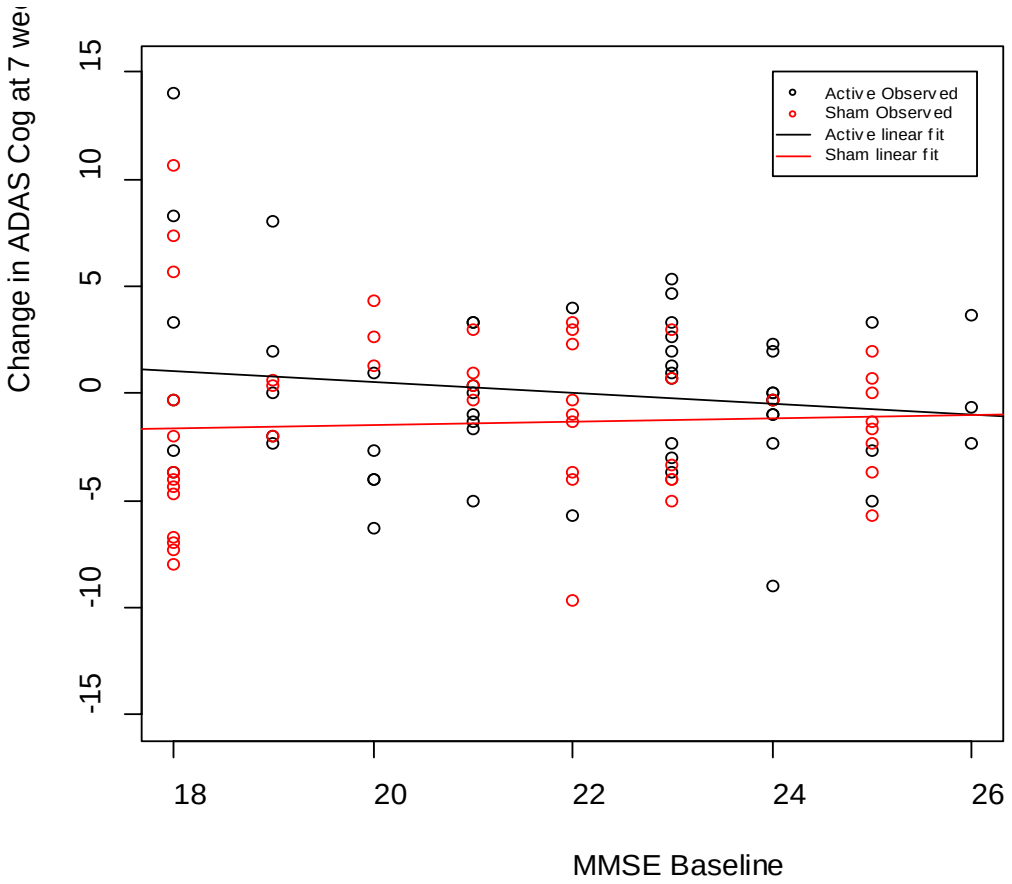
Subgroup x Treatment Interaction p = 0.03

A number of different post-hoc subgroups might be identified that could separate the population into responders and non-responders.

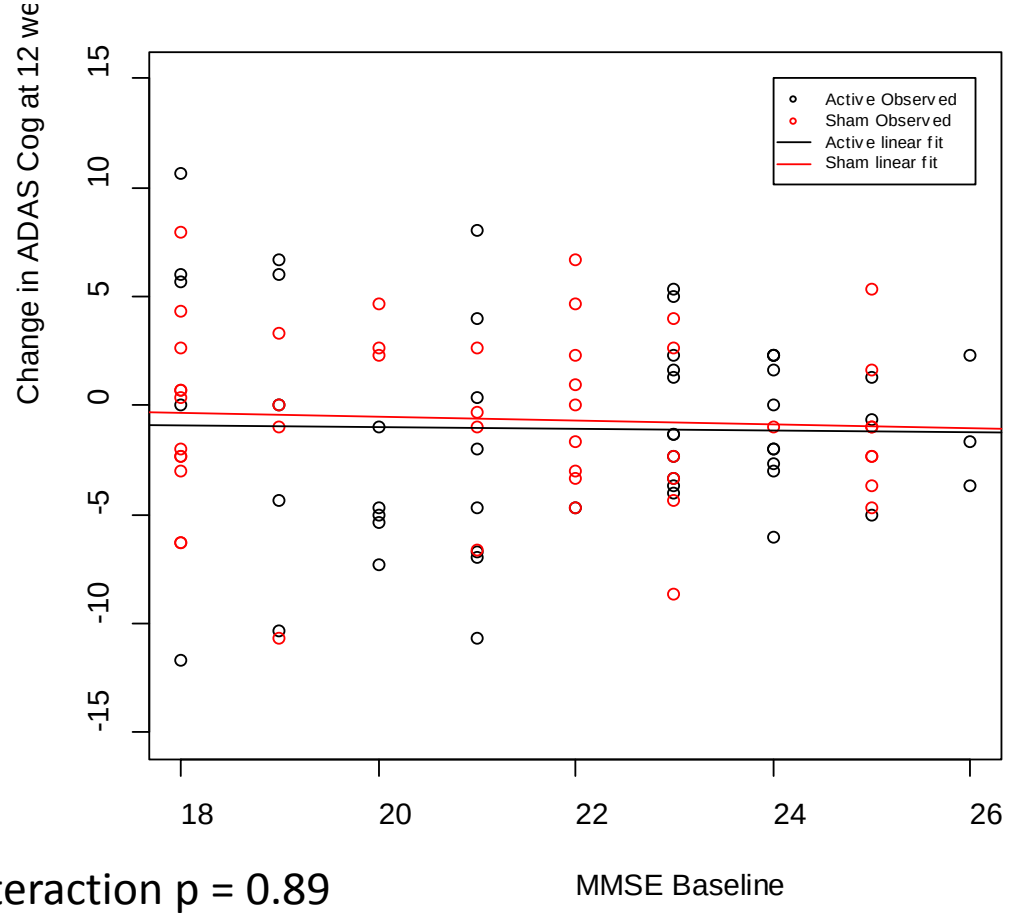
Data-driven Hypotheses based on Subgroups

- A similar IA does not appear to occur with supplemental studies.
- If there is a real IA between AD severity and treatment group (neuroAD vs. sham), we would expect to see a similar IA between baseline MMSE and treatment group.
- However, an IA with baseline MMSE was not seen…

Interaction between Treatment Group and Baseline MMSE (also AD severity)

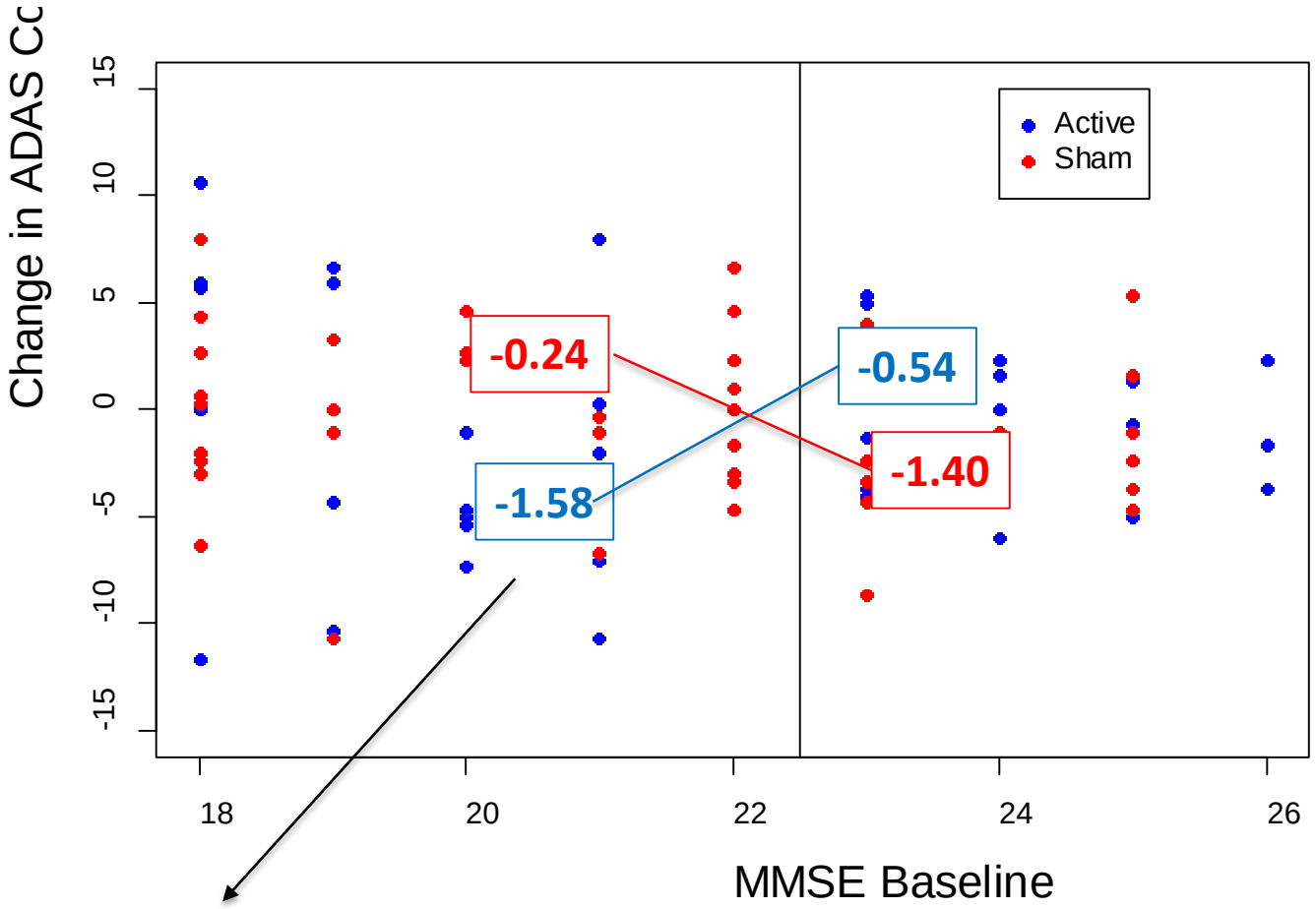


Interaction p = 0.31



Interaction p = 0.89

Post-Hoc Baseline MMSE Cut-point <= 22.5



Subgroup x Treatment Interaction p = 0.24

Trt diff in subgroup p = 0.37

Post-hoc Cut-point

- Literature does not support that a device like neuroAD works better in baseline ADASCog<30.
  - Lee et al. and Ito et al. support *prognostic* effect of MMSE or intensity of AD.
  - If baseline ADAS-Cog measured in US pivotal study *did* mean AD severity, it would be correlated with MMSE. Its correlation (-0.45) is much lower than correlations seen in Ito et al. meta-analysis (-0.90)
- A patient with lower disease severity will be expected to benefit more from *any* treatment or care.

Principles for Co-Development of an In-vitro Companion Diagnostic Device with a Therapeutic Product

- "When the use of an IVD companion diagnostic is essential for the safe and effective use of the therapeutic product and its use is part of the instructions for use of the therapeutic product, FDA recommends that, whenever possible, the candidate IVD companion diagnostic be validated as part of the major efficacy trial(s)."
- The set of clinical samples used to design an IVD and establish the clinical decision point(s) and assay cutoff(s) is referred to as the "training set." Testing should be conducted with a second set of independent clinical samples (i.e., the "validation set") and with the final IVD design to validate the IVD and determine whether the assay cutoffs correlate with clinical outcome.
- For IVD companion diagnostics, the validation sample set is generally made up of samples from subjects screened for enrollment into the major efficacy clinical trial(s) that is intended to support efficacy claims for the therapeutic product. For this reason, IVD design and assay cutoffs should be established before the IVD is applied to these samples.

Companion Diagnostic Guidance: External validation of the cut-point

- *If changes are made to the IVD based on results obtained with the clinical samples from the major efficacy trial(s) (e.g., changing the cutoff to include all those who responded in the trial), then what would otherwise have been the validation set effectively becomes a new training set for the modified IVD.*
- *The modified IVD likely could not receive marketing authorization as an IVD companion diagnostic without further studies, as it will likely not select the same population represented in the major efficacy trial(s).*
- *While it may seem logical to use the trial specimens to assure concordance between the two versions of the test, there is no assurance as to whether the same concordance would be obtained with a different set of samples. The new IVD design may be established with a set of procured clinical samples similar to the subjects in the trial or samples from earlier investigational trials.*

Exploratory or Hypothesis-Generating Analyses*

- Investigators make mistakes not because they perform exploratory analyses, but when they represent such findings as the primary results of the trial. It is essential to acknowledge the hypothetical nature of exploratory findings and recognize that the usual calculation of type I errors may be incorrect, especially when the data themselves suggest the hypothesis test. As a general rule, the same data should not be used both to generate a new hypothesis and to test it.

*Excerpted from Piantadosi's *Clinical Trials, A Methodologic Perspecitve*, page 320, Wiley 1997

Subset Analyses are Error Prone*

One of the easiest ways for the analysis of a clinical trial to follow an inappropriate direction occurs when investigators emphasize the findings from a particular subset of patients, especially when the results are different from the overall findings (i.e. an analysis including all randomized patients). These interesting results may be found in a particular subset of patients after an extensive search that is not based on any *a priori* biological hypothesis. Other times, an accidental observation in a subset may suggest a difference, which is then tested and found to be "statistically significant". If the investigators have prejudices or reasons from outside the trial to believe the findings, these circumstances could lead to a fairly firmly held belief in the validity of the results. Unfortunately, the potential for error in this scenario is high.

* Excerpted from Piantadosi's *Clinical Trials, A Methodologic Perspecitve*, pages 321, Wiley 1997

Data-driven Hypotheses based on Post-hoc Subgroups

- A decision based on a post-hoc hypothesis test is subject to error due to potentially spurious finding. In particular:
  - The usual calculation of type I errors may be incorrect due to using the data to generate a (new) hypothesis and to test it.
  - An accidental observation in a subgroup may suggest a 'statistically significant difference .... If the investigators have prejudices … to believe the findings, these circumstances could lead to a fairly firmly held belief in the validity of the results. Unfortunately, the potential for error in this scenario is high." (Piantadosi, 1997)
- The baseline ADAS-Cog <= 30 is not necessarily "accidental"…
  - A similar IA does not appear to occur with supplemental studies.
  - The cut-point is from a measurement instrument, with measurement error

Internal Validation of the cut-point

- Internal validation largely upheld a treatment effect in lower baseline ADAS-Cog values, but with smaller magnitude effect.
- Despite the internal validation using different “training” and “validation” sets for each iteration of the procedure, the two sets originate from the same data set.
- External validation is most reliable because the external data set is (assumed to be) completely independent of the original study.

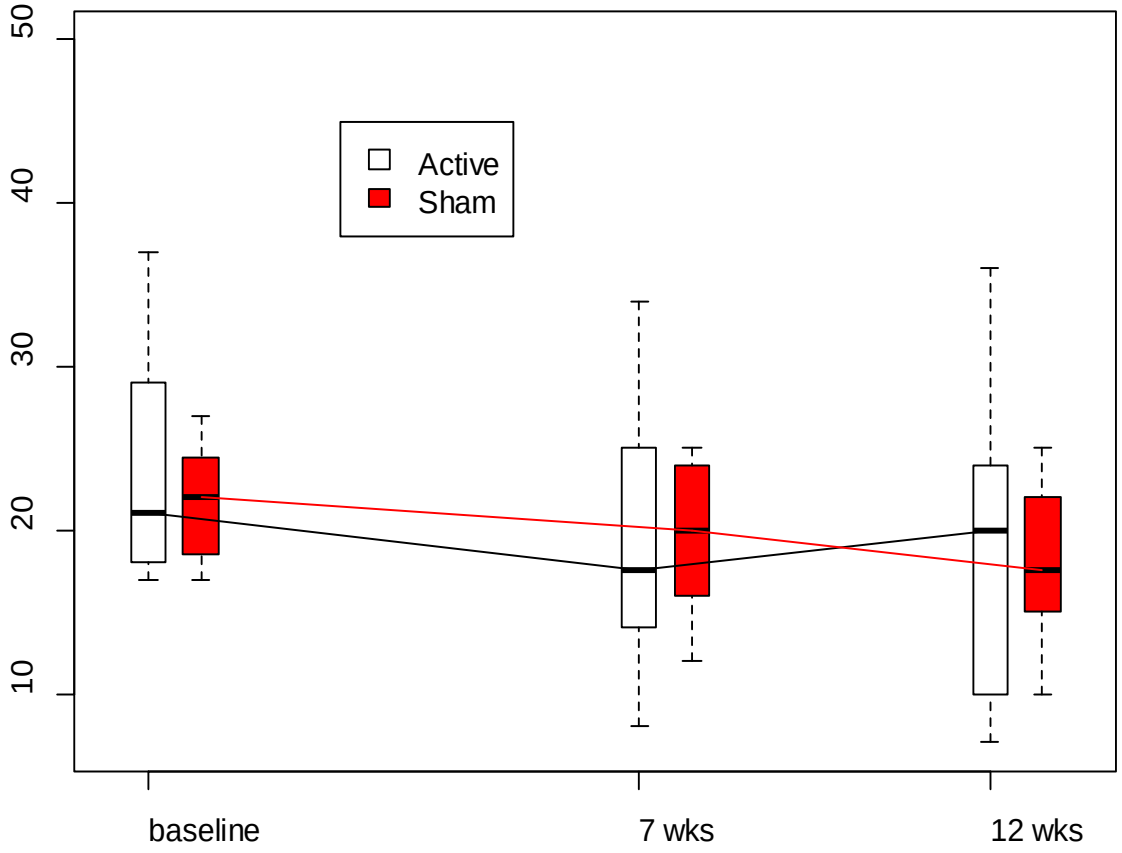
Internal validation of the cut-point

- Internal validation is often used when an external validation is not possible.
- The procedure repeatedly divides the study data into training and validation sets such that for each training set, a “best” cut-point is obtained, and then tested on the validation set by estimating the treatment effect in the “best” subgroup definition.
- Adjusted treatment effect estimate in subgroup that corrects for bias due to a potential random high (selection using observed results).

Internal validation of the cut-point: Bias-adjusted estimates of treatment effect

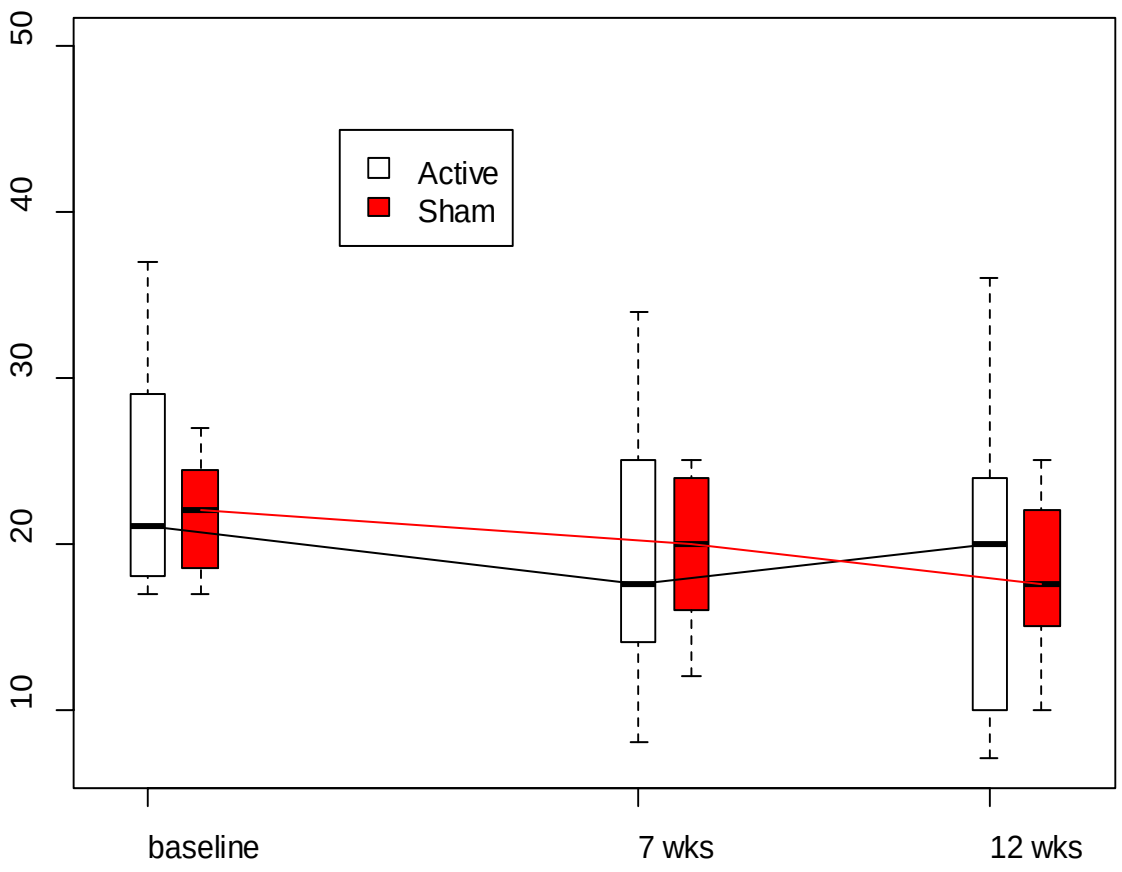
	Adjusted estimate of treatment effect	Observed Result
In subgroup	-1.33 (-2.33, -0.43)	-1.61 (-3.39, 0.17)
Enhanced difference of subgroup over the entire population	-0.76 (-1.46, -0.06)	(-1.61 – (-0.42)) = -1.19

Change in ADAS-Cog by Assessment Time
Korea Pilot Study

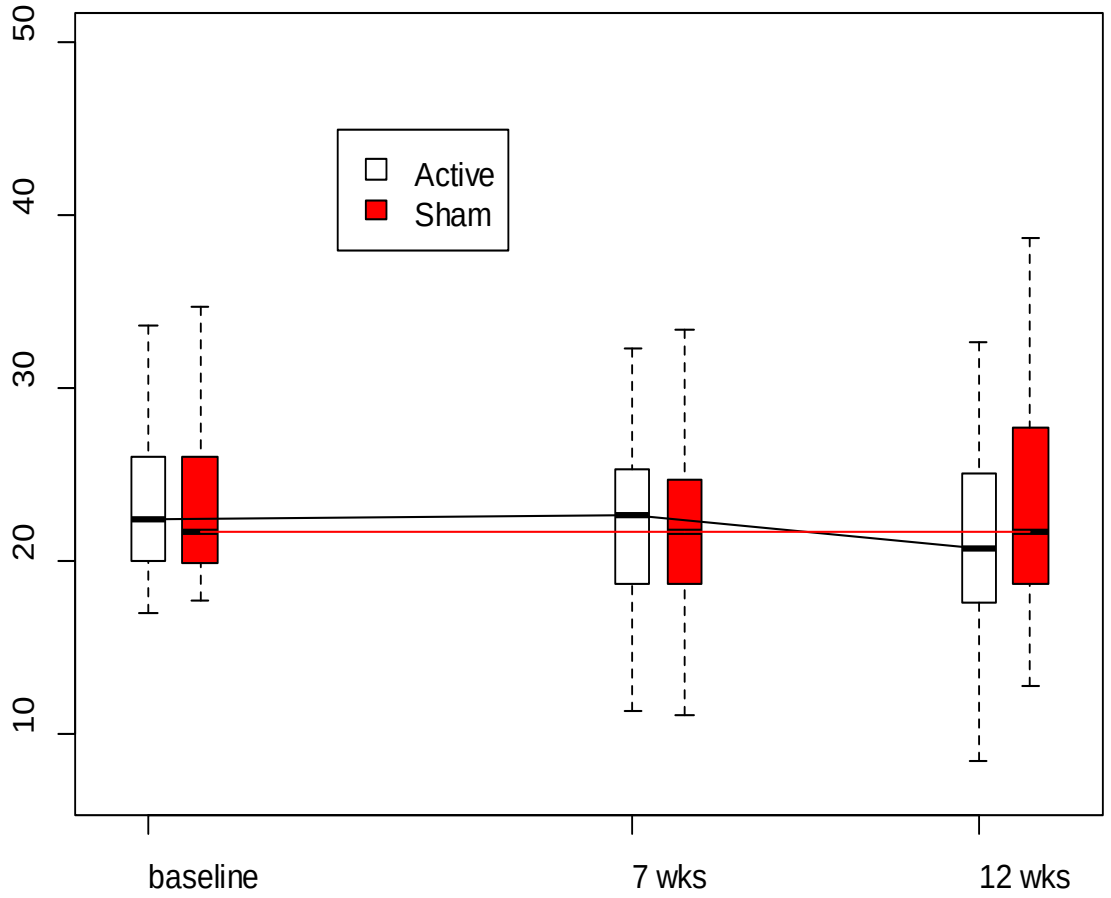


	Change at 7 weeks	Change at 12 weeks
Active (n=18)	-4.28	-5.39
Sham (n=9)	-1.75	-2.88
Difference	**-2.53** (-6.33, 1.27)	**-2.51** (-6.39, 1.35)
p-value	0.18	0.19

Korea Pilot vs. US Full Cohort – By Assessment Time



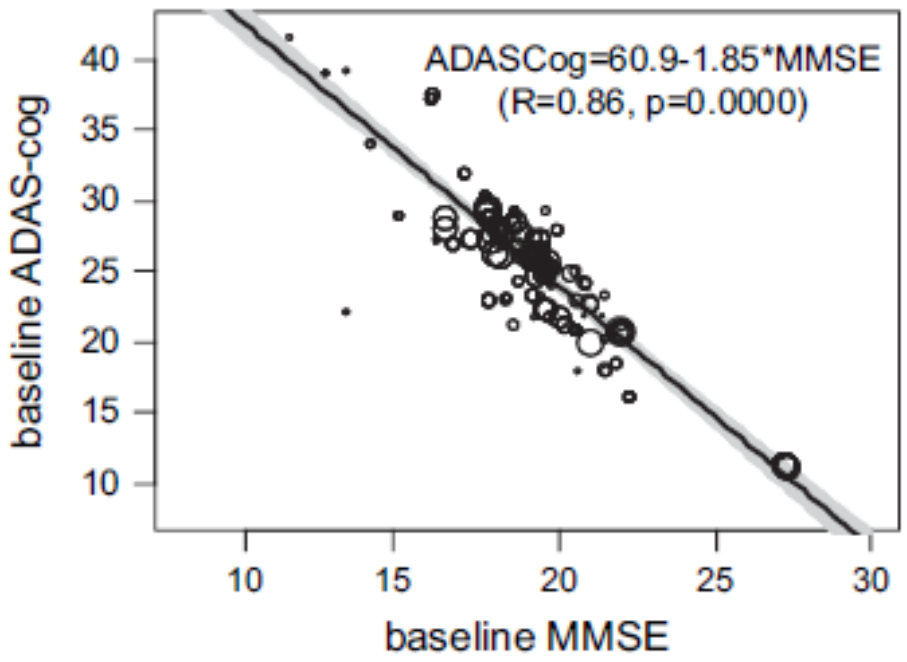
Korea Pilot

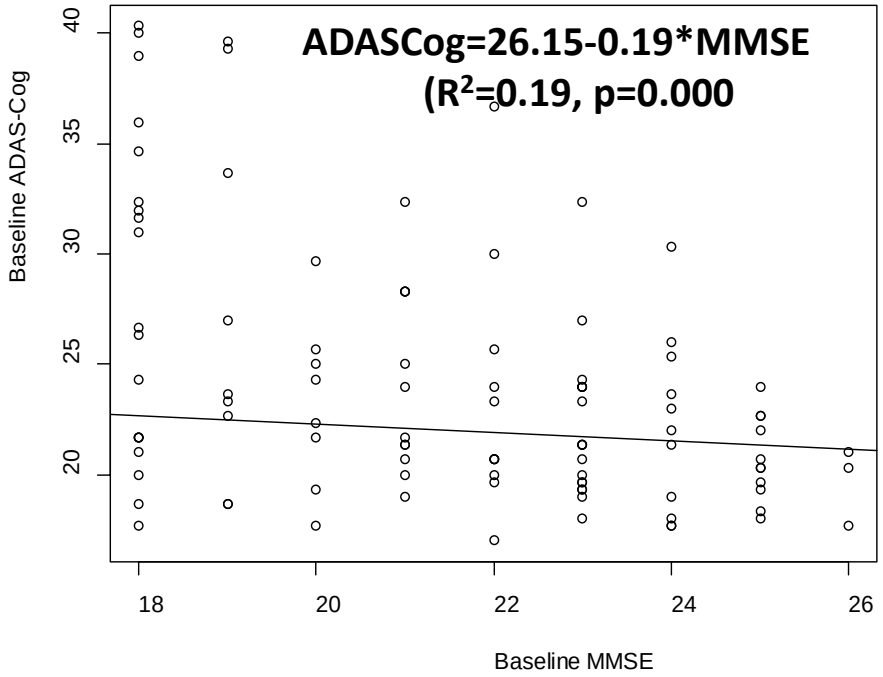


FDA Bayesian Hierarchical Model for Adjusted Subgroup-specific Treatment Effect Estimate

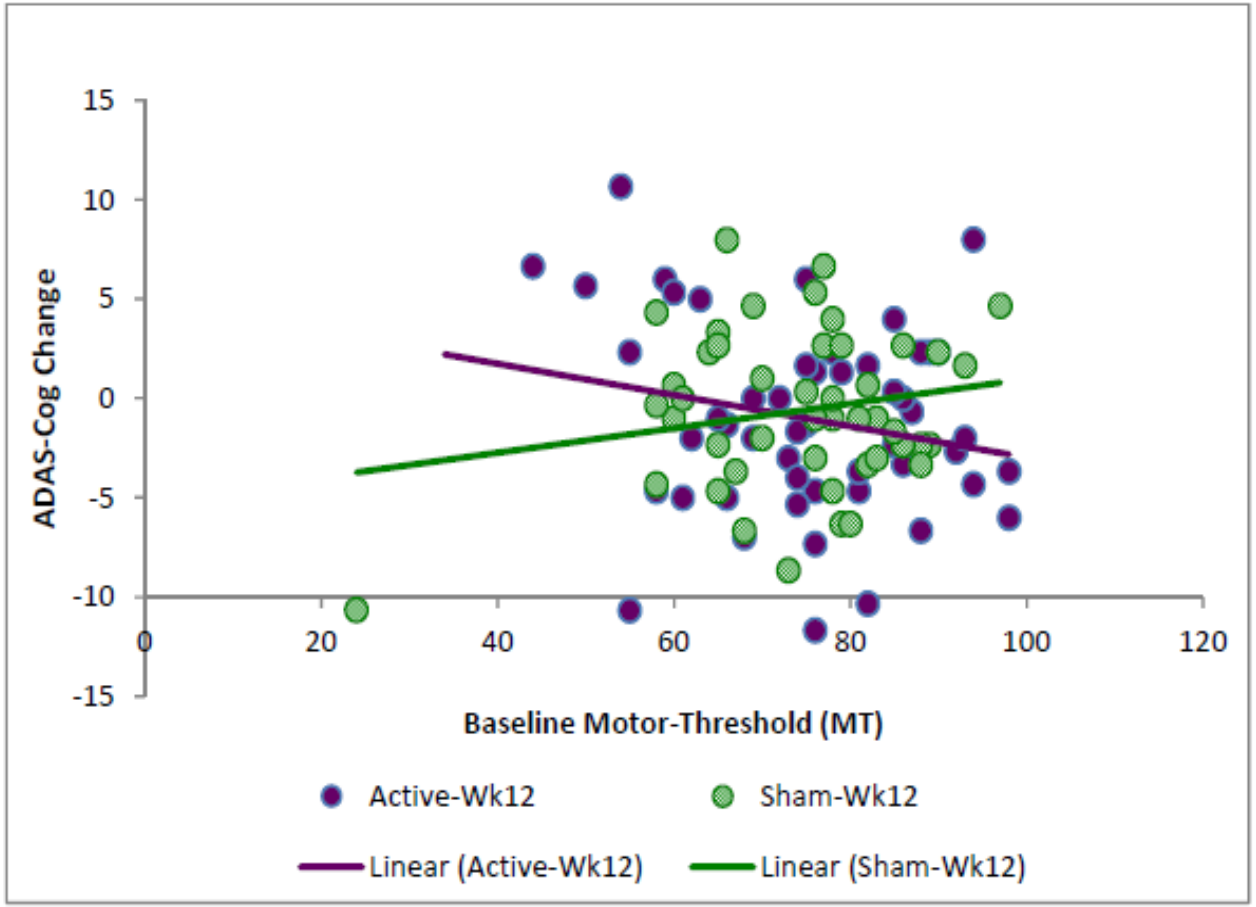
	Non-hierarchial subgroup specific estimate (95% CI)	Hierarchical model subgroup specific estimate (95% CI)
Treatment difference (ADAS Cog baseline<=30)	-1.61 (-3.35, 0.13)	**-1.39** (-3.26, 0.46)
Treatment difference (ADAS Cog baseline > 30)	6.61 (1.96, 11.26)	**5.34** (0.05, 9.93)

Ito et al. (2010) vs US pivotal Study





Lack of Relationship between MT and Treatment Difference on ADAS-Cog Change



External validation of the cut-point: baseline ADAS-Cog <= 30

	Korea Studies Combined (25 Active; 18 Sham)	Korea Studies + Assaf3 Studies Combined (33 Active; 23 Sham)	All RCTs combined (Korea+Assaf3+Italy) (38 Active; 24 Sham)
In subgroup	–2.12 (–4.66, 0.43)	–2.57 (–5.05, -0.09)	–2.32 (–4.69, 0.045)
Enhanced difference of subgroup over the entire population	–2.12 – (–2.4) = +0.28	-2.57 – (-2.95) = +0.38	-2.32– (-2.60) = +0.28

How To Proceed?

❌ 1. Should we pool data across studies?

- Studies are exchangeable, not poolable

❌ 2. Use only "good" subgroup from Study 1?

- Capitalizes on potentially spurious post-hoc finding in "good" subgroup, and discards data.

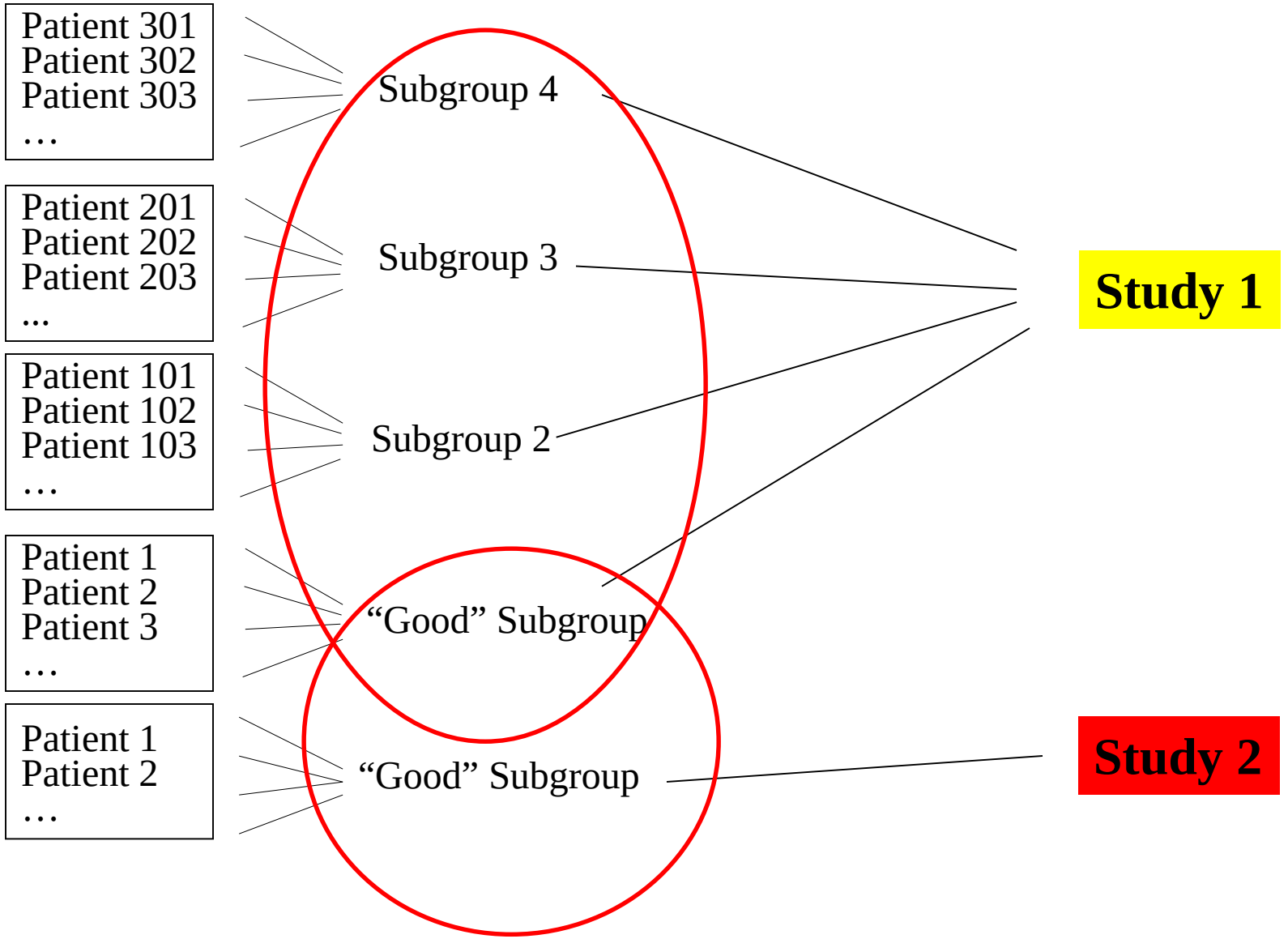
〰 3. Analyze second study alone?

- Wastes information

✔ 4. Allow Study 2 to borrow from *all* of Study 1.

- With multiplicity adjustment to down-weight potentially spurious results.

Hierarchical Structure



FDA's Bayesian Hierarchical Model (HM)

- Even if exchangeability holds, all subjects should be used in analysis
- FDA hierarchical model assumed exchangeable studies and exchangeable subgroups within studies.
  - HM implements borrowing across studies as well as borrowing across subgroups within studies
  - Subgroup-specific treatment effect estimates may be exaggerated because the subgroup is identified based on a (high) favorable result.
  - HM "shrinks" these estimates to more reasonable values that might align better with results from pre-specified subgroup analyses.

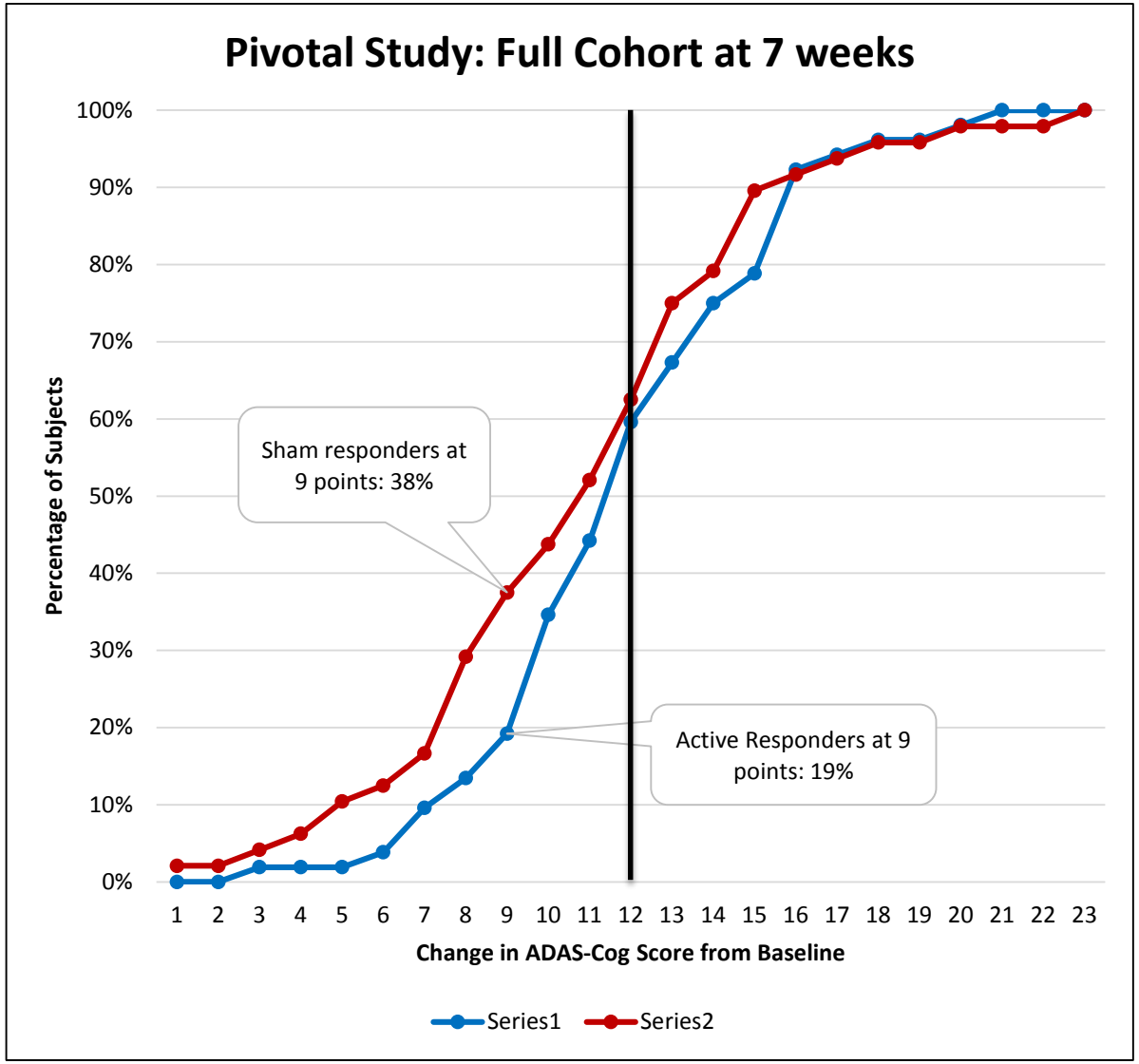
Sponsor's Meta Analysis vs. FDA Bayesian Hierarchical Model

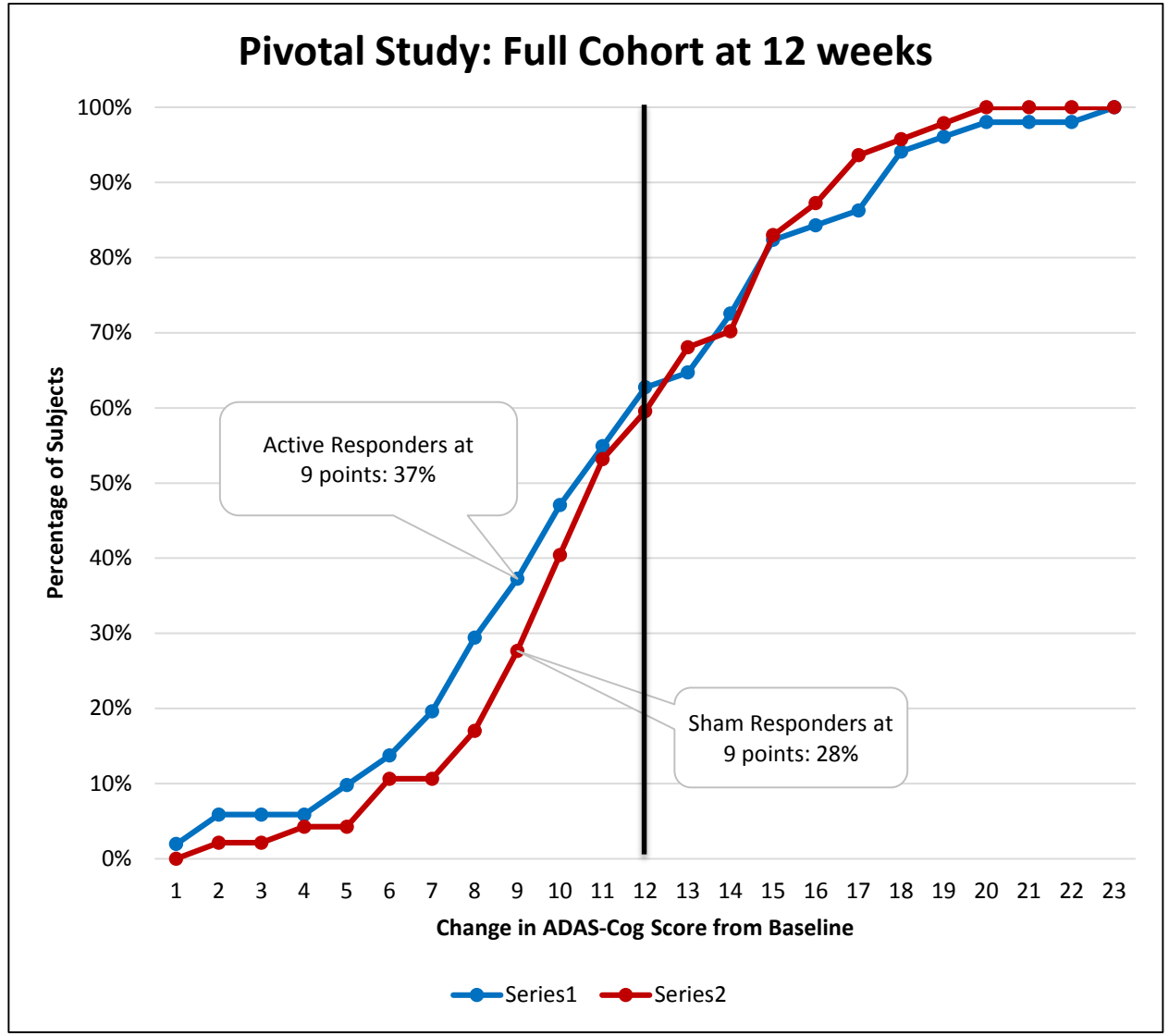
	US Estimate in baseline ADAS-Cog <=30	US Estimate in baseline ADAS-Cog > 30	Common or Pooled Estimate in baseline ADAS-Cog <=30	Common or Pooled Estimate in baseline ADAS-Cog > 30
Sponsor's meta analysis (WMD)	-1.61 (-3.36, 0.14)	NA	-1.66 (-3.03, -0.29)	NA
Observed result	-1.61 (-3.35, 0.13)	+6.61 (1.96, 11.26)	-1.90 (-3.41, -0.40)	+2.72 (-2.27, 7.70)
FDA's hierarchical model analysis	-1.26 (-2.93, 0.45)	+2.15 (-1.95, +6.17)	-1.61 (-4.67, 0.88)	+1.73 (-2.91, 6.39)

What would be a way to proceed with Korea-2 as confirmatory?

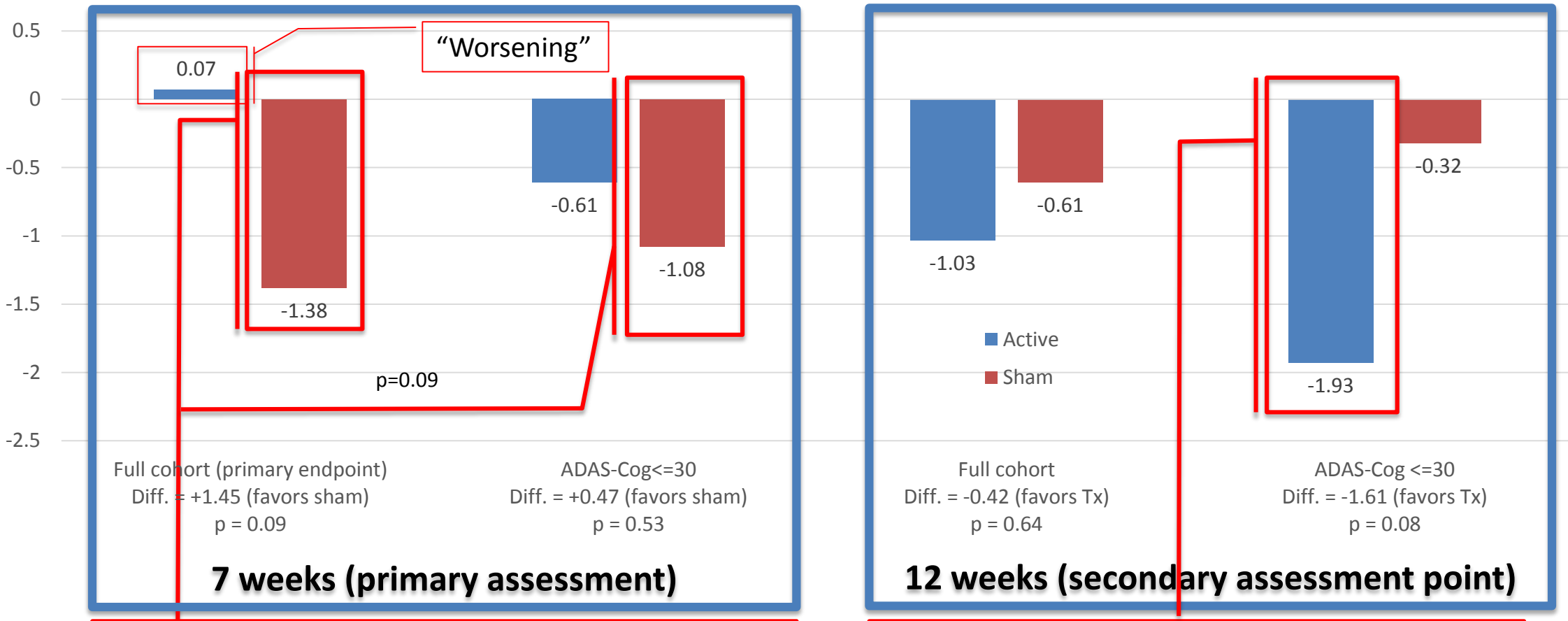
- If we assume that <= 30 cutpoint is reliable, how would this subgroup be “confirmed” in a new dataset (e.g., Korea-2)?
- What would the confirmed result be?
  - We could use the Bayesian HM to (re)estimate the treatment difference in the Korea-2, adjusting for previous US study (all subjects).
  - Studies are assumed exchangeable, subgroups within studies are exchangeable (>30 subgroup in the Korea-2 study is imputed).
  - Adjusted treatment effect estimate: -1.51 (-3.85, 0.59). The posterior probability that the mean difference is < 0 = 0.93

Cumulative Proportion of Responders





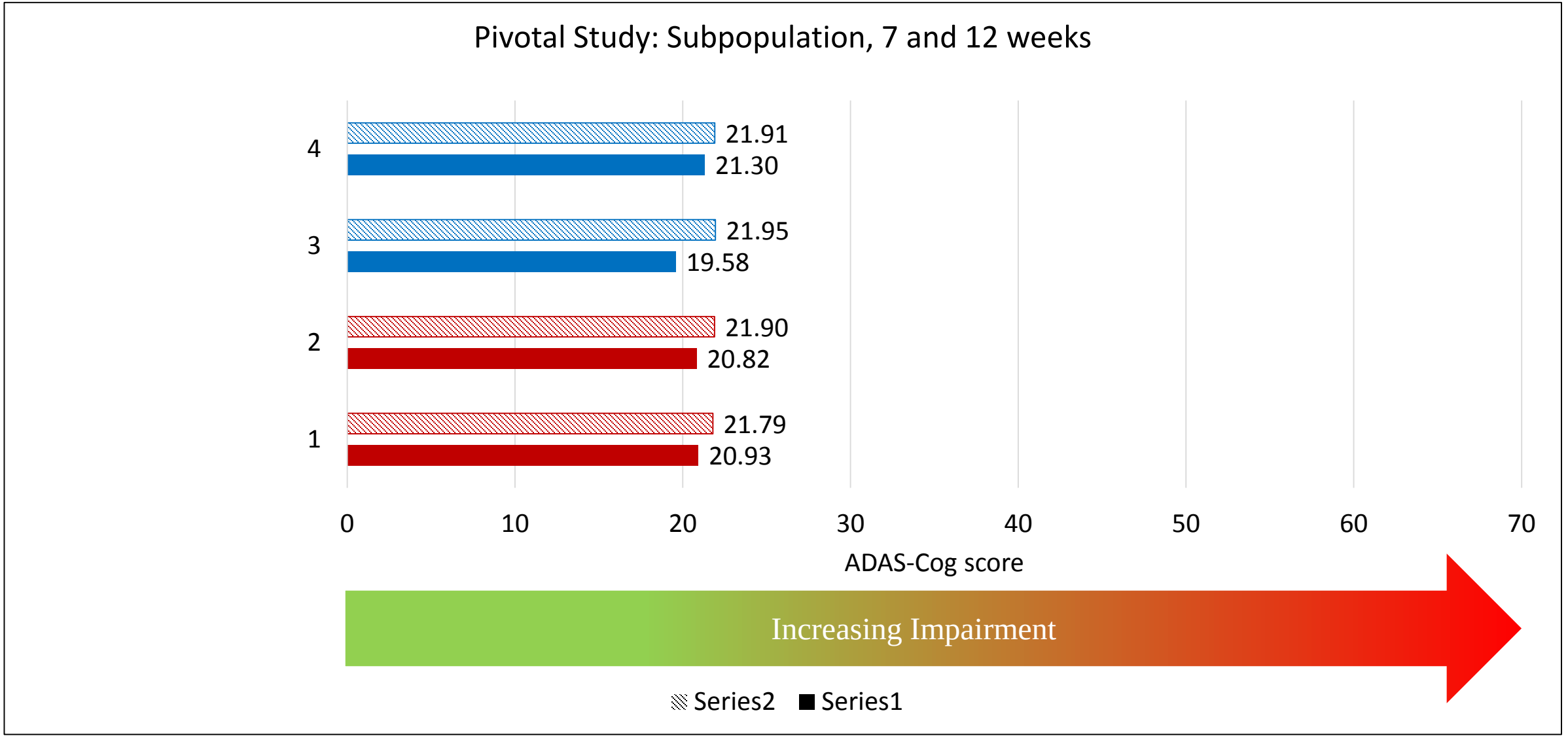
Uncertainty in Pivotal Results – ADAS-Cog

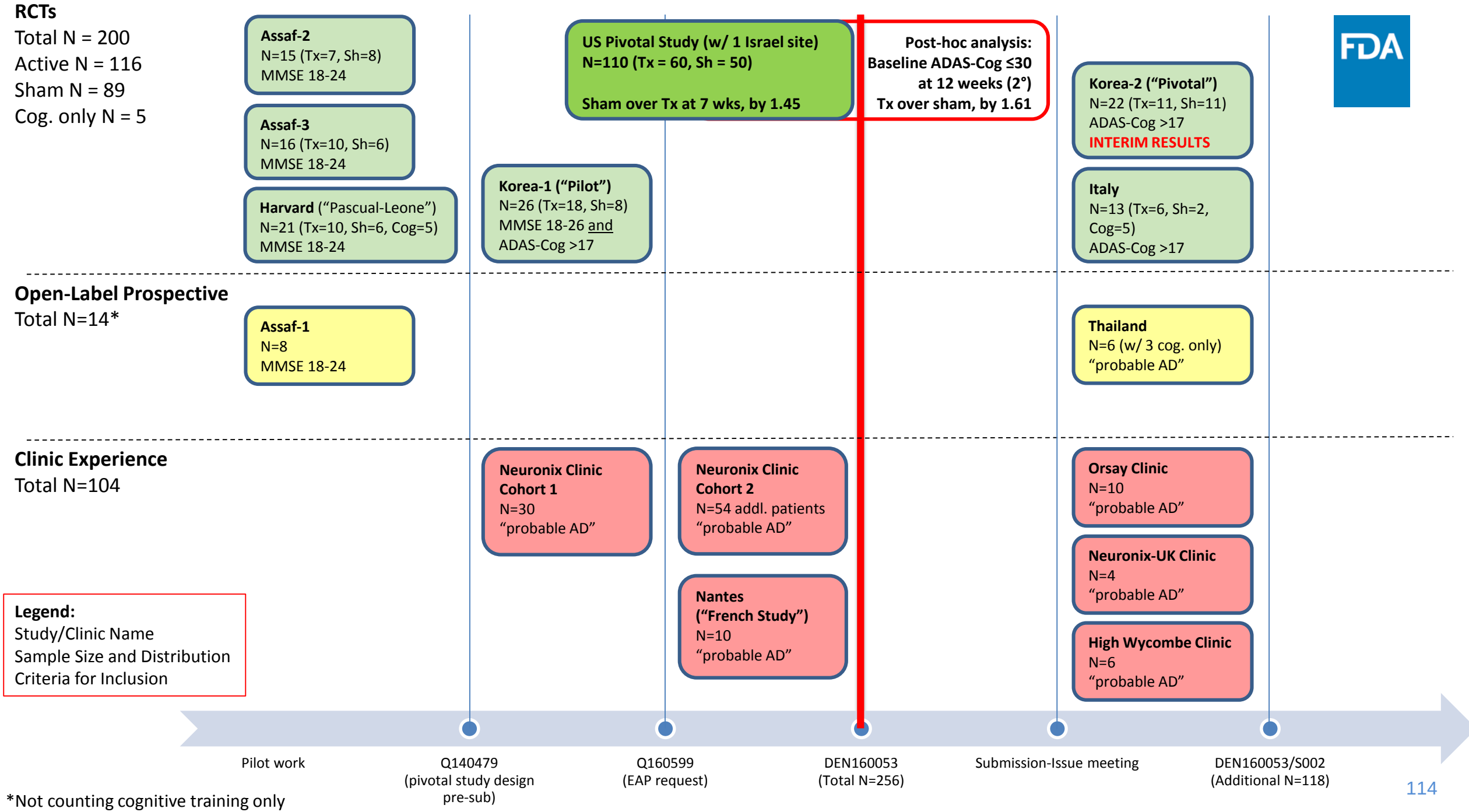


Sham numerically improves more than the treatment group in both the overall population and selected subpopulation.

Reversal of direction of treatment difference at 12wk endpoint was not seen in other studies; subgroup was determined post-hoc

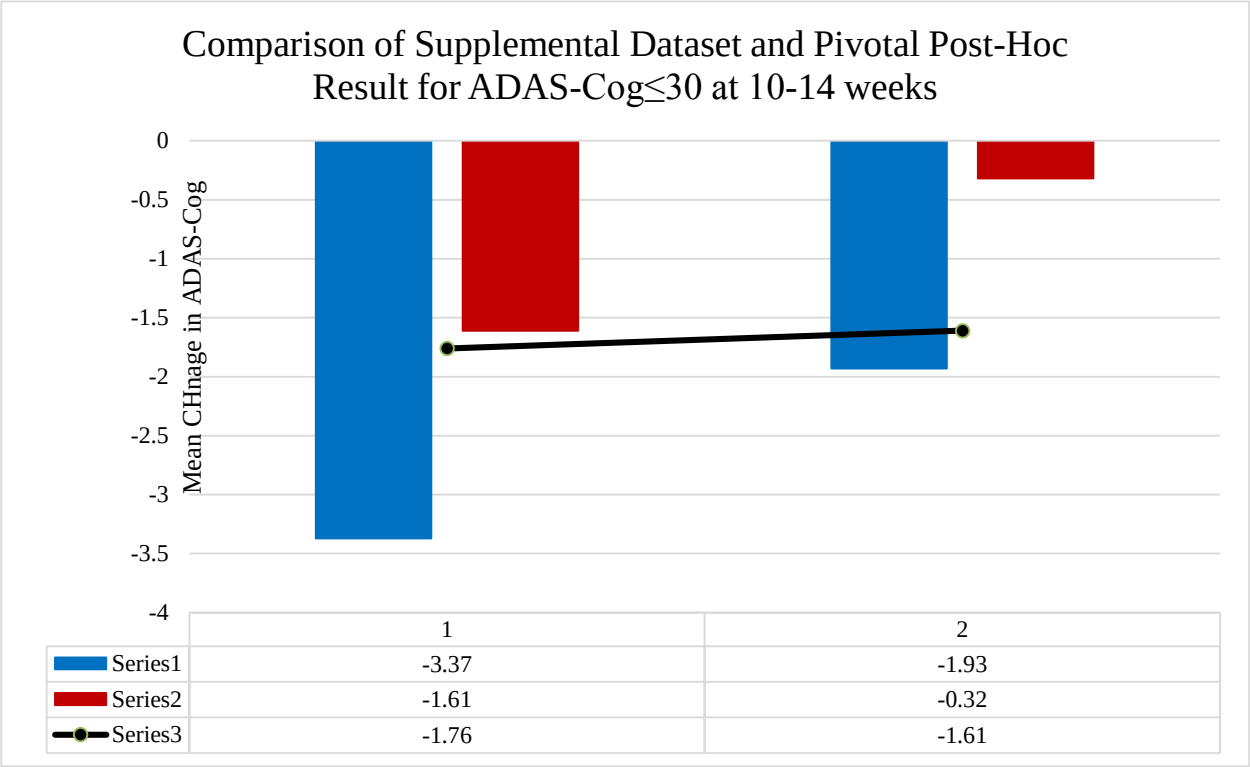
Summary Graph of Pivotal Study ADAS-Cog Results





*Not counting cognitive training only

Supplemental Investigations to Verify Pivotal Post-Hoc Analysis



Stakeholder Input
Sponsor's Physician Survey – MCID on ADAS-Cog

Sponsor Claims Physicians Support that 1 point (or less) is Clinically Meaningful

- Sponsor conducted an unsolicited survey between February 23 – March 16, 2018 of neurologists and psychiatrists to assess what they considered clinically meaningful.

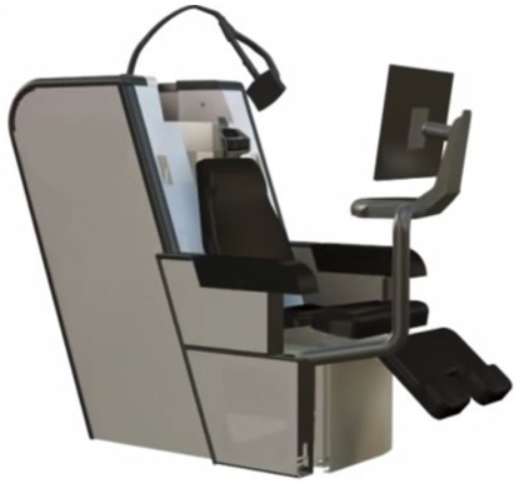
- The sponsor concludes that "Nearly half of physicians consider *at least a 1 point improvement (or less, so long as there is no deterioration) in ADAS-Cog Score* clinically meaningful following 3 months of treatment; even more find this threshold clinically meaningful when there is also a 0.5 point improvement in ADCS-CGI-C."

- However, both the methods and results raise concerns about validity.

Survey Concern – Methods (1)

neuroAD (Blinded) Product Profile

- A new, non-invasive, medical device treatment for Alzheimer's Disease, typically administered in combination with pharmacotherapy.
- Treatment is administered at the clinic for one hour per day, 5 times/week, over 6 weeks.
- It utilizes two modalities concurrently combined:
  1. Neuro-navigated focused Transcranial Magnetic Stimulation (TMS) is used to stimulate targeted areas of the brain responsible for various cognitive functions that have been impaired by Alzheimer's disease.
  2. Tailored Cognitive Training is used to target those same areas of the brain while they are being magnetically stimulated.
- It has minimal-to-no side effects, and provides cognitive and functional improvement.
- Results from a pivotal study of patients with a baseline ADAS-Cog score of 17-30 (mild-to-moderate Alzheimer's Disease):



ADAS-Cog Score	Treatment Group subjects experienced a mean improvement in score at 12 weeks of -2.11. The same subpopulation in the placebo group reported a mean change of -0.32 (between groups difference of -1.79 favoring the treatment, statistically significant).
ADCS-CGI-C Score	Treatment Group subjects reported a lower mean CGI-C score than the placebo group with difference between groups of -0.45 (favoring treatment) at week 12 (nearing statistical significance, p=0.07). CGI-C distribution reached statistical significance.

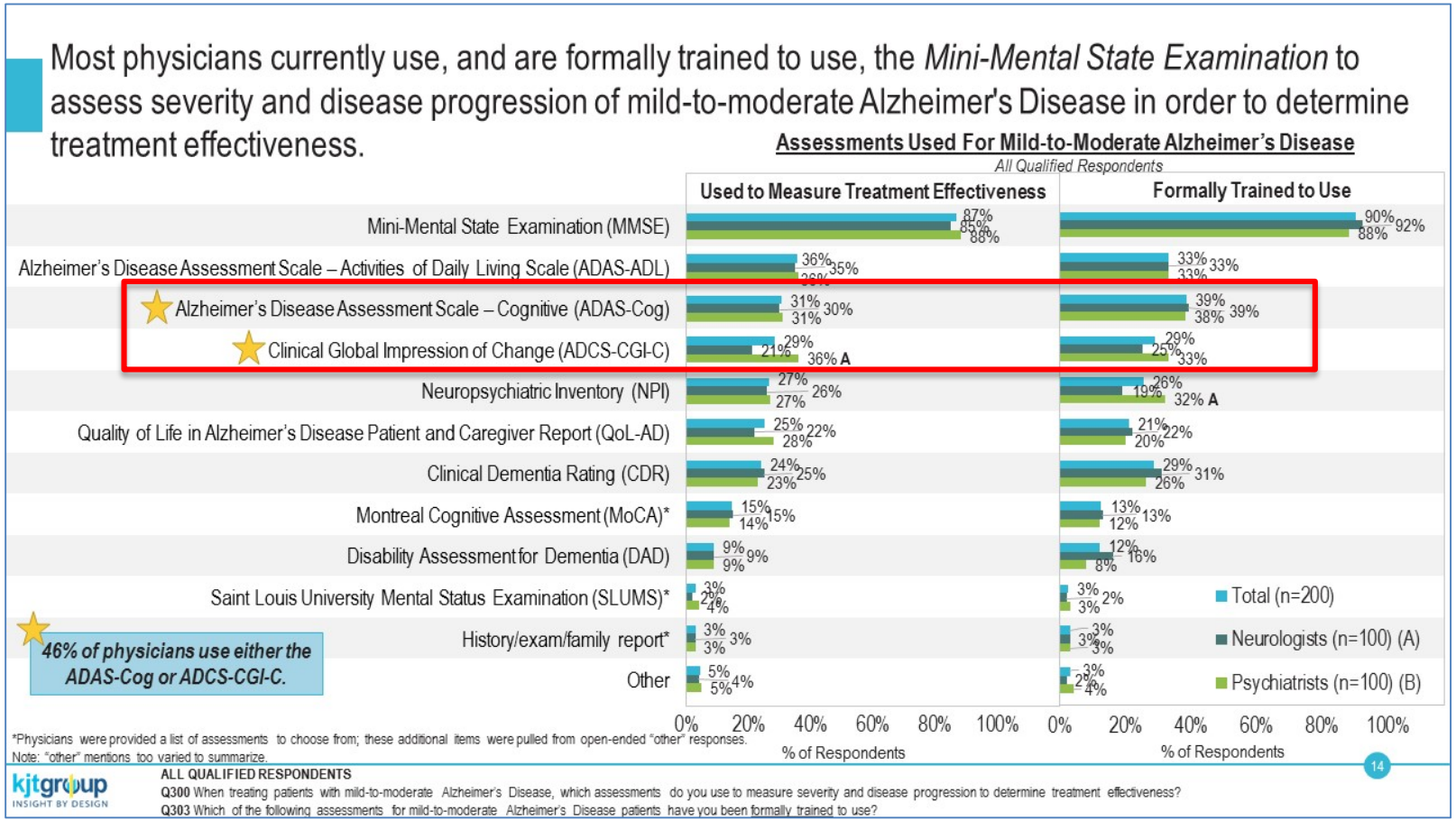
kjtgroup INSIGHT BY DESIGN

19

The sponsor overstated the "benefits" of the device:

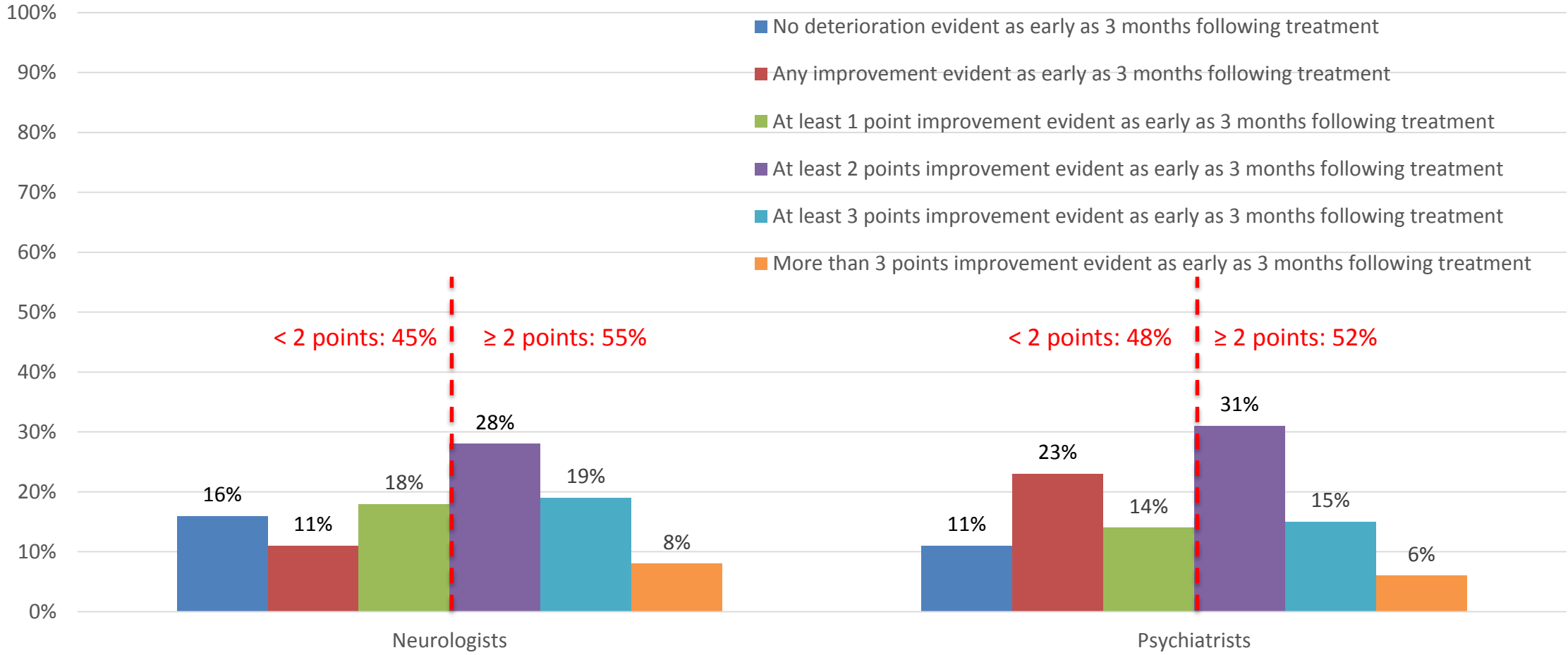
- Told survey applicants that the device provides cognitive and functional improvement
- The results presented are the per-protocol (PP) population and not Primary Efficacy (PE) results, which excluded those with major protocol deviations (7/8 excluded subjects were poor-performing active subjects)

Survey Concern – Methods (2)



- Only 31% of Respondents *use* the ADAS-Cog for which they are providing an assessment of *clinical* meaningfulness
- Only 29% use the CGI-C which is used in follow-up question to support a smaller change in ADAS-Cog as MCID

Survey Question Q310: "Now, again thinking from your perspective, what minimal degree of improvement in the ADAS-Cog score would you consider clinically meaningful when evaluating the effect of a new treatment administered on top of Cholinesterase inhibitors to your patients with mild-to-moderate Alzheimer's Disease."



Survey Concern – Interpretation of Results

Sponsor's Interpretation

Nearly half of physicians consider *at least a 1 point improvement (or less, so long as there is no deterioration) in ADAS-Cog Score* clinically meaningful following 3 months of treatment; even more find this threshold clinically meaningful when there is also a 0.5 point improvement in ADCS-CGI-C.

Team Interpretation

More than half of physicians considered *at least 2 points or greater* on the ADAS-Cog score to be clinically meaningful following 3 months of treatment

- Without an additional CGI-C improvement, most physicians (30%) responded that at least 2 points improvement on the ADAS-Cog is clinically meaningful.
- With the additional CGI-C of 0.5 points improvement, the same is true though by a smaller margin (28%), and is driven by psychiatrists

Note: Pivotal Study data never reached CGI-C of 0.5

Network of Experts – MCID on ADAS-Cog (March-April, 2018)

Q. Assuming negligible risk from an intervention, what is the smallest demonstrated change in the ADAS-Cog that you would consider clinically meaningful enough to try the intervention on a patient with Alzheimer's disease?

- #1: 4-5 points per year, 2 points would not be clinically meaningful
- #2: 4 points, plus other measures
- #3: Saw 1-3 points in the literature with placebo worsening and treatment staying the same over 6 months suggesting a disease modifying therapy (not studied in this investigation)

“Freedom From Deterioration”

Note: Pivotal study was designed to support therapy indication not disease modification

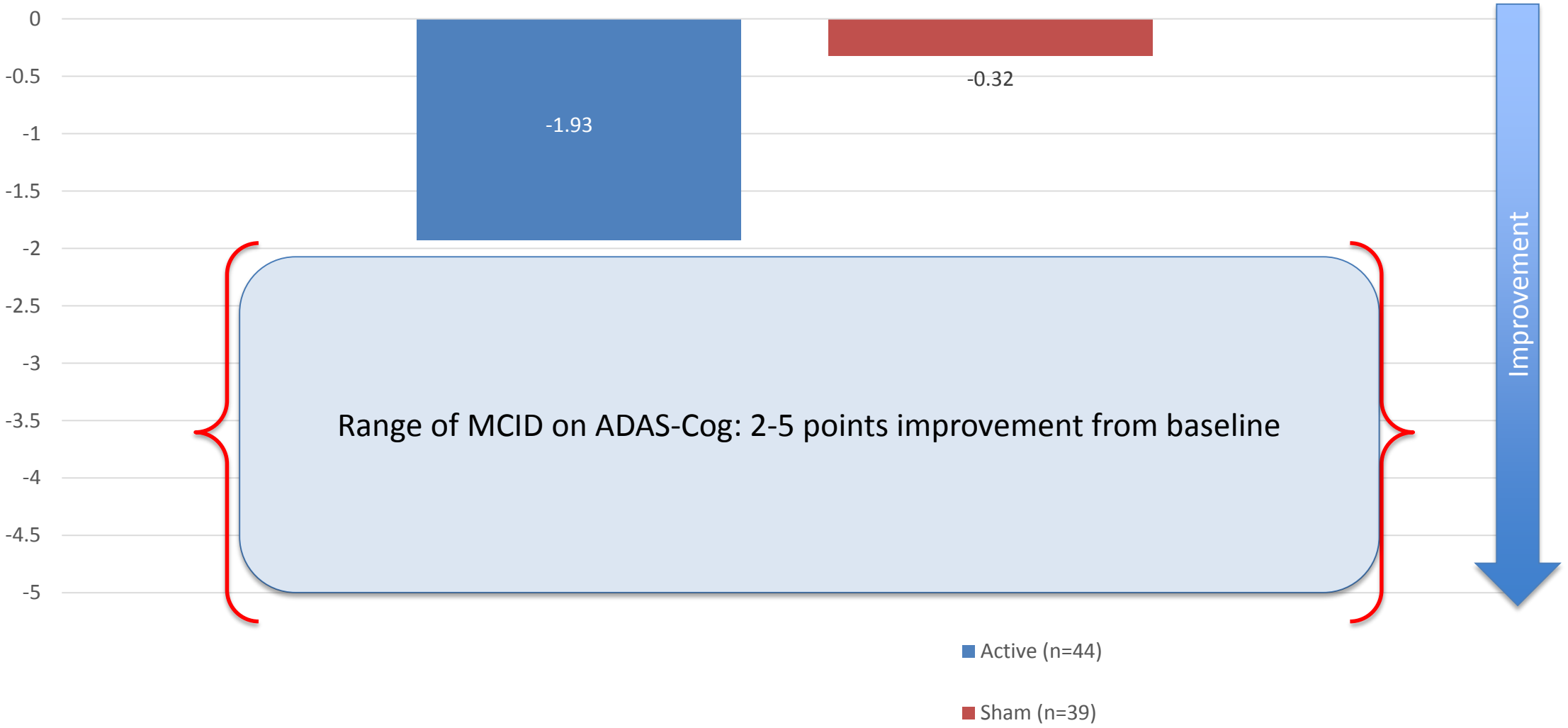
<u>CGI-C results do not support this claim</u>

- Neither secondary endpoints nor responder analyses were intended for labeling claims (Q140479).
- CGI-C results show fewer Active subjects in “worsening” categories (5-7) at 12 weeks, which is upheld in other studies. However, the proportion showing “improvement” on CGI-C (1-3) is identical across treatment groups (regardless of subgroup) – 28%.
  - **If we are to accept a claim of freedom from deterioration, then we should also accept a claim of no improvement over sham.**

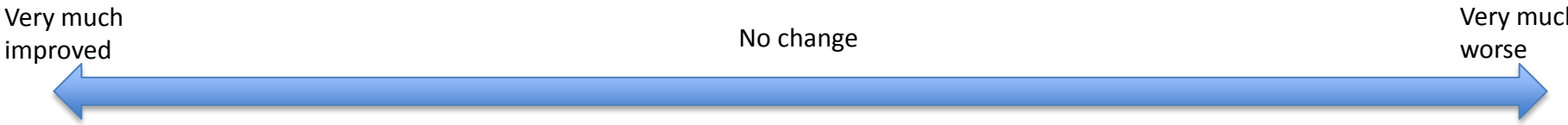
“Freedom From Deterioration”

- Pivotal study was not designed to demonstrate freedom from deterioration
  - Freedom of Deterioration is a novel claim that was not prespecified, discussed during the conduct of the study, nor has been reported in the literature
  - Length was too short (7 week primary, 12 week secondary)
  - Assessments may not be appropriate depending on scope of “deterioration”

MCID In Context:
"Best"* Pivotal Results – Post-hoc Analysis
ADAS-Cog at 12 weeks, Mean diff = -1.61

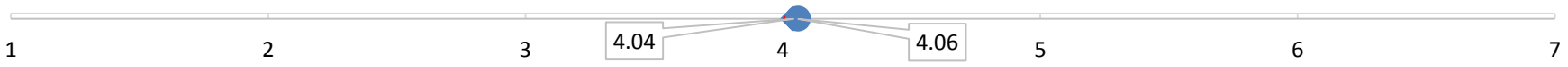


Summary of Pivotal Results – CGI-C

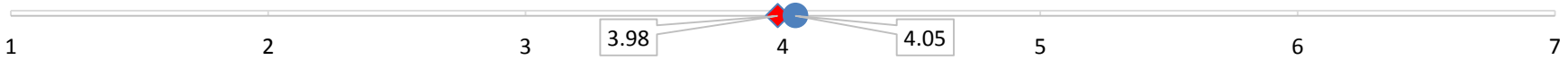


7-week Assessment

Full Cohort



Post-hoc analysis: Baseline ADAS-Cog ≤30



12-week Assessment

Full Cohort

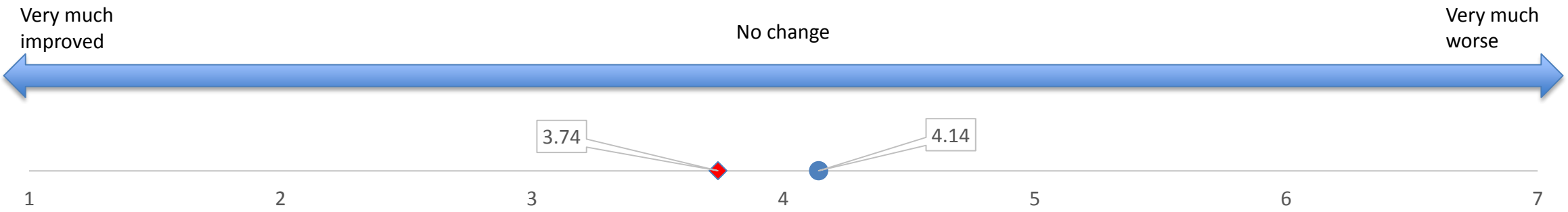
Post-hoc analysis: Baseline ADAS-Cog ≤30

Best-case difference: -0.4

"Best" Pivotal Results – Post-hoc Analysis
CGI-C AT 12 Weeks

Review team considered CGI-C as a way to provide context to -1.61 difference in ADAS-Cog. CGI-C was a secondary endpoint. The results below are limited to same ADAS-Cog<=30 subgroup.

This "best-case" CGI-C result shows difference between groups is **0.4 points** in favor of treatment – this is too small to provide support for clinical benefit



Red diamond is active
Blue circle is sham

Table 1: Combined Sample for ADAS-Cog Assessment

Study	n	Subjects with Change in ADAS-Cog Data at 6-10 weeks		Subjects with Change in ADAS-Cog Data at 10-14 Weeks	
		Active	Sham	Active	Sham
Assaf 1 (Israel, Open label)	8	8	0	0	0
Assaf 2 (Israel, Double Blind, controlled)	15	7	8	0	0
Assaf 3 (Israel, Double Blind, controlled)	16	10	6	10	6
Beth-Israel study, (Harvard, USA, Double Blind, controlled)	21	10	6	10	0
Korean pilot study (Korea, Double Blind, controlled)	26	18	8	18	8
Korean pivotal study (Korea, controlled)	22	11	11	11	11
Nantes clinic (France)	10	10	0	0	0
Italian study (Italy, controlled)	13	6	2	6	2
NeuroCare (Neuronix clinic, Israel, commercial cases)*	84	59	0	10	0

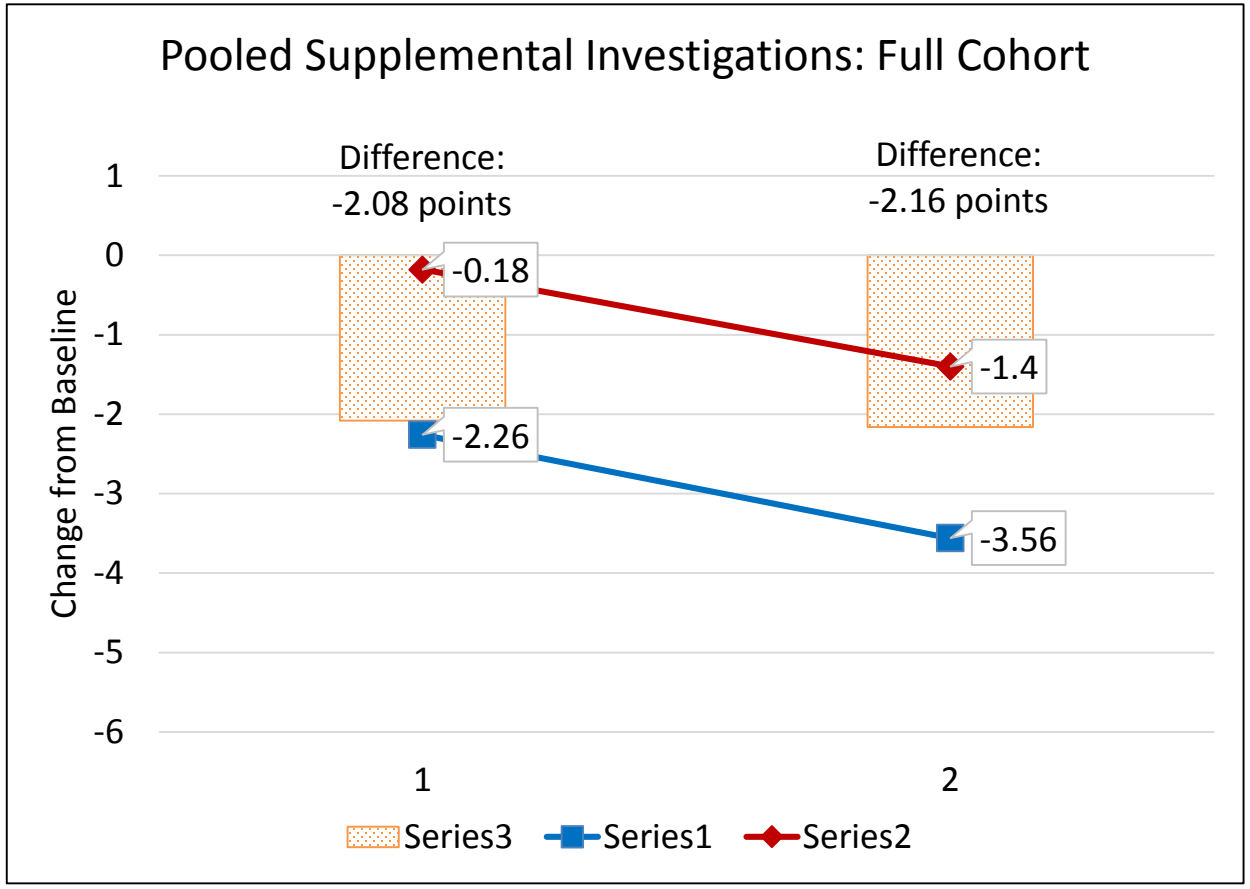
Clinical materials for reply to FDA deficiency letter 19-Sep-

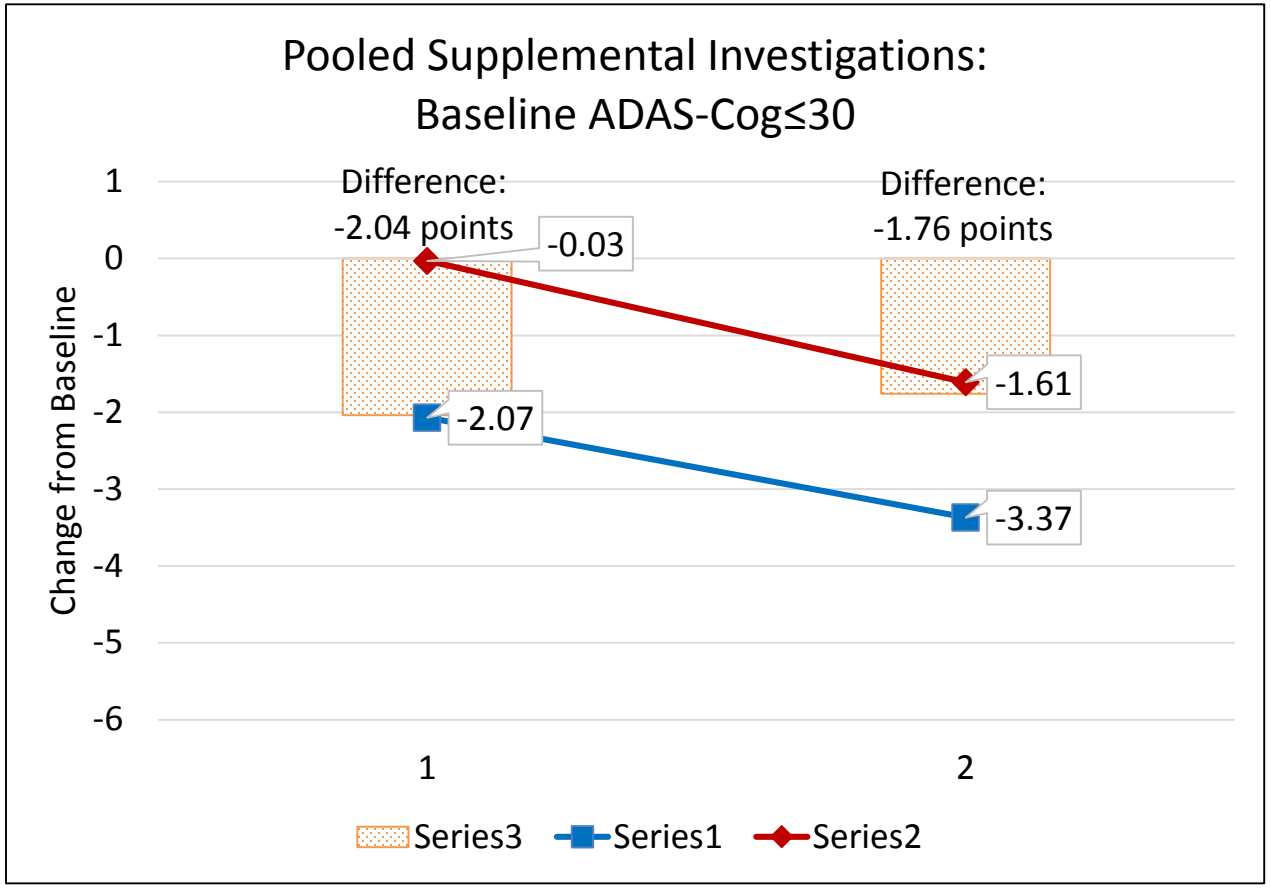
Study	n	Subjects with Change in ADAS-Cog Data at 6-10 weeks		Subjects with Change in ADAS-Cog Data at 10-14 Weeks	
		Active	Sham	Active	Sham
Total	**215[1]**	**139**	**41**	**65**	**27**

[1] Of the 215 subjects included in the combined analysis 97 subjects provide new data not previously submitted to FDA.

Pooled Supplemental Data: Effectiveness Results by Subgroup

Both active and sham groups show similar trends in improvement over time, both for the full cohort and subpopulation. However, the full cohort performs better at both time points.

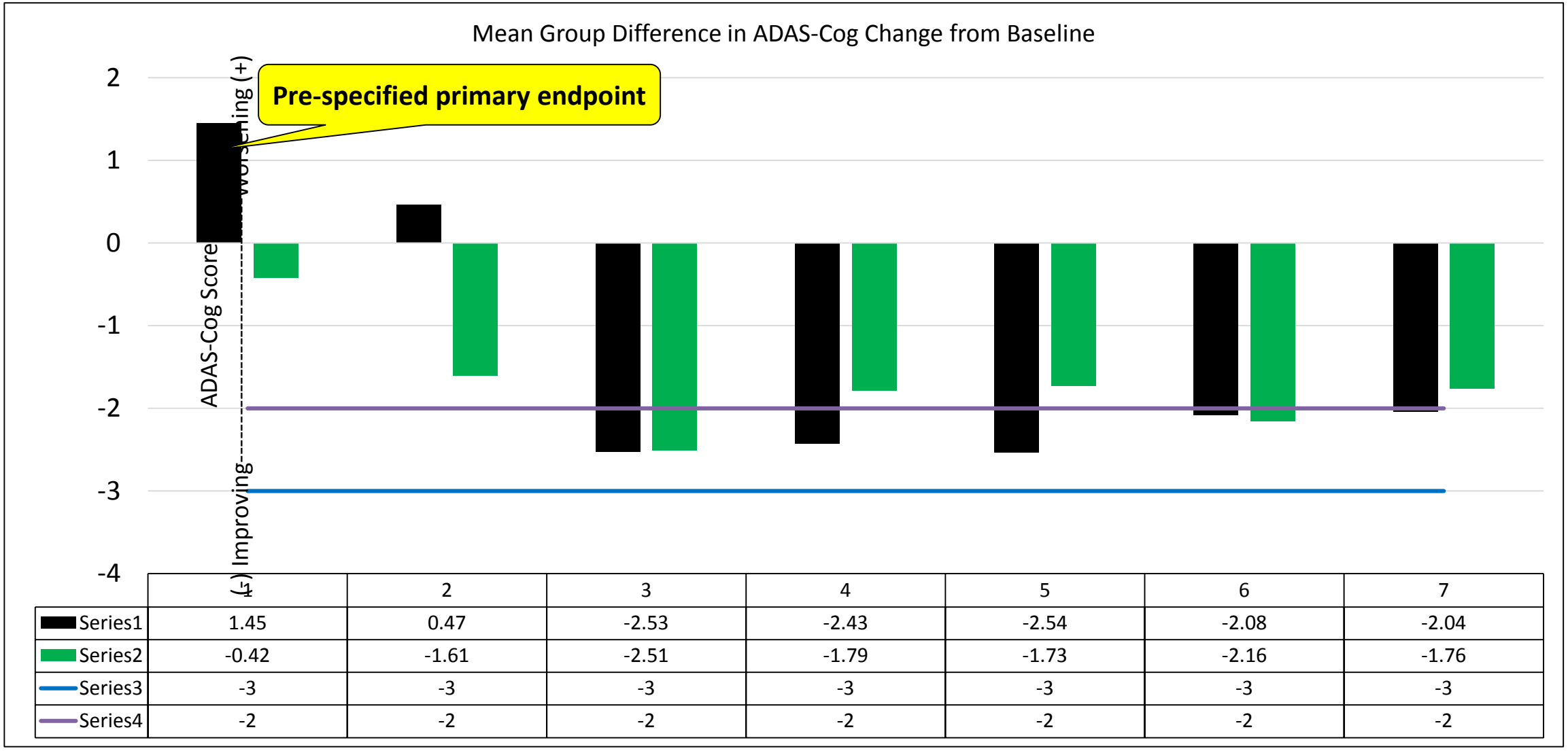




Uncertainty with Supplemental Investigations: Application to US Population

- Look at OUS extrapolation guidance
- Not prespecified
- Small sample sizes
- Other interactions happening? Like Motor Threshold?
- Dr. Thompson will discuss more…

MCID In Context – All Data Sources



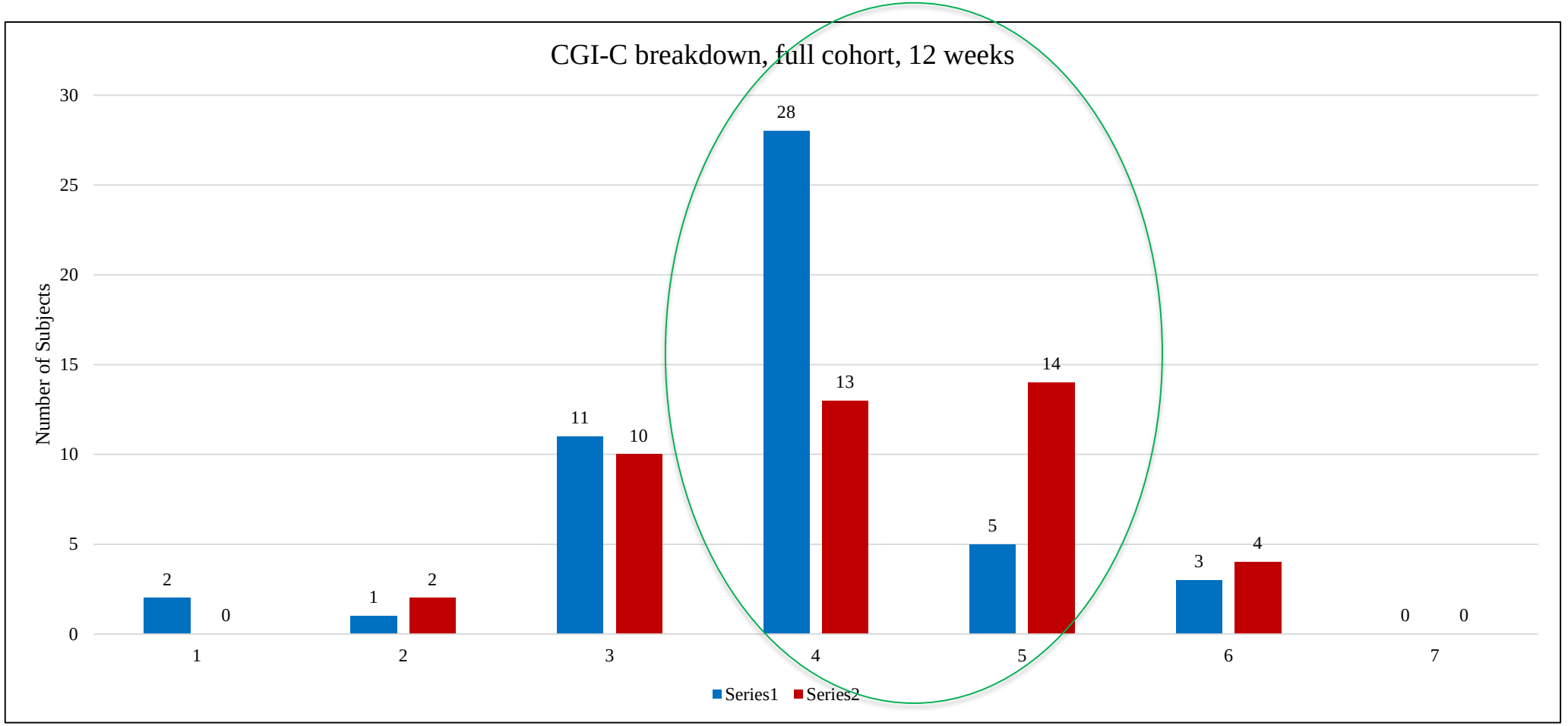
	1	2	3	4	5	6	7
Series1	1.45	0.47	-2.53	-2.43	-2.54	-2.08	-2.04
Series2	-0.42	-1.61	-2.51	-1.79	-1.73	-2.16	-1.76
Series3	-3	-3	-3	-3	-3	-3	-3
Series4	-2	-2	-2	-2	-2	-2	-2

Valid Scientific Evidence

- 21 CFR 860.7 (c)(2): Valid scientific evidence is evidence from well-controlled investigations, partially controlled studies, studies and objective trials without matched controls, well-documented case histories conducted by qualified experts, and reports of significant human experience with a marketed device, from which it can fairly and responsibly be concluded by qualified experts that there is reasonable assurance of the safety and effectiveness of a device under its conditions of use.
- It is unclear whether the neuroAD meets this definition

Backup or delete

Distribution of CGI-C at 12 weeks



Korea Studies may not verify US Pivotal Post-Hoc Subgroup

- Small sample sizes
- Korea-1, included Baseline ADAS-Cog ≤30 and >30
- Korea-2, limited to Baseline ADAS-Cog ≤30
  - "[t]he present results suggest that rTMS-COG represents a useful adjuvant therapy with cholinesterase inhibitors, particularly during the **mild stage** of AD." (Lee et al., 2016)
  - **Stages of AD were defined by MMSE**: "The participants were categorized into mild [Mini-Mental State Examination (MMSE) score=21–26] and moderate (MMSE score=18–20) AD groups" (Lee et al., 2016)

TMS Parameter Selection

- Intensity was patient-specific, based on motor threshold (MT)
  - The intensity when a motor reaction occurs
  - Measured daily in the active group
  - Intensity was determined by the area being stimulated, varying between 90-110% of the motor threshold
- Other stimulation parameters are fixed
- Parameters of neuroAD magnetic stimulation pulses are consistent with marketed systems

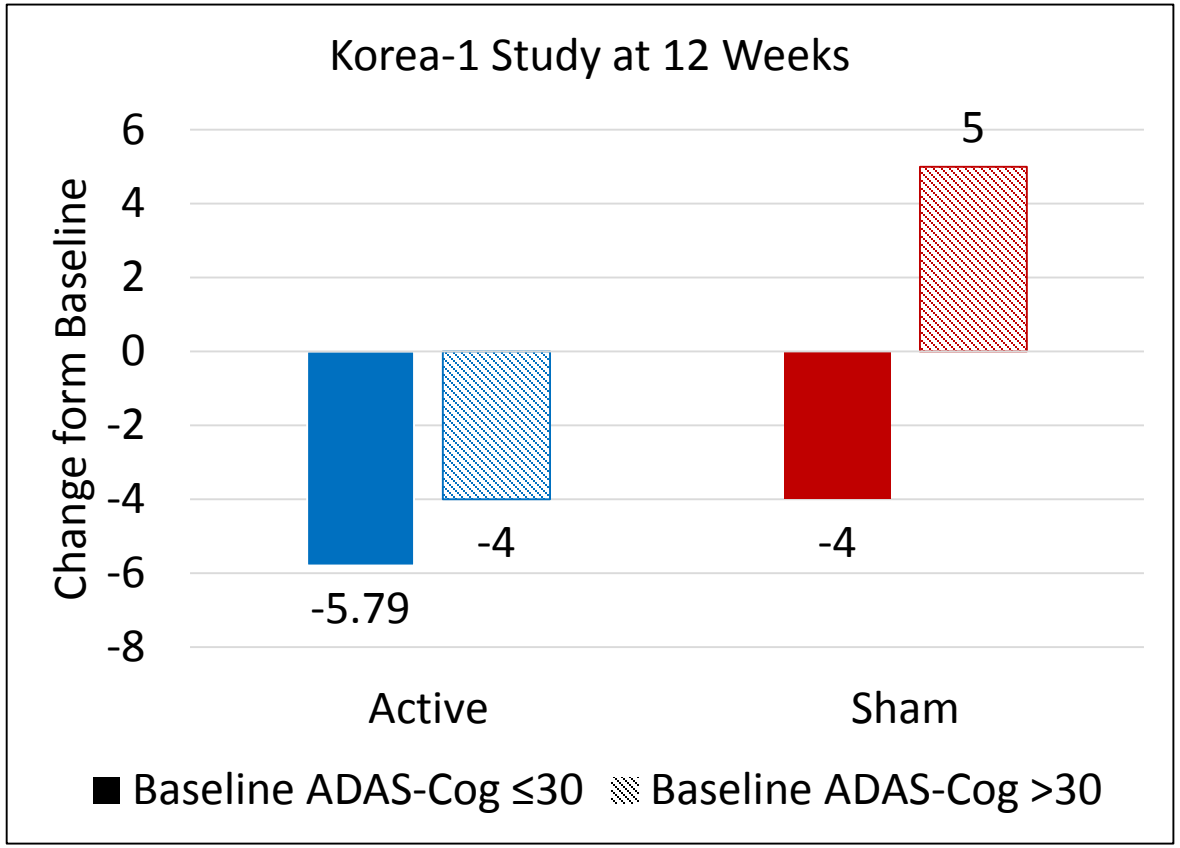
Cognitive Training

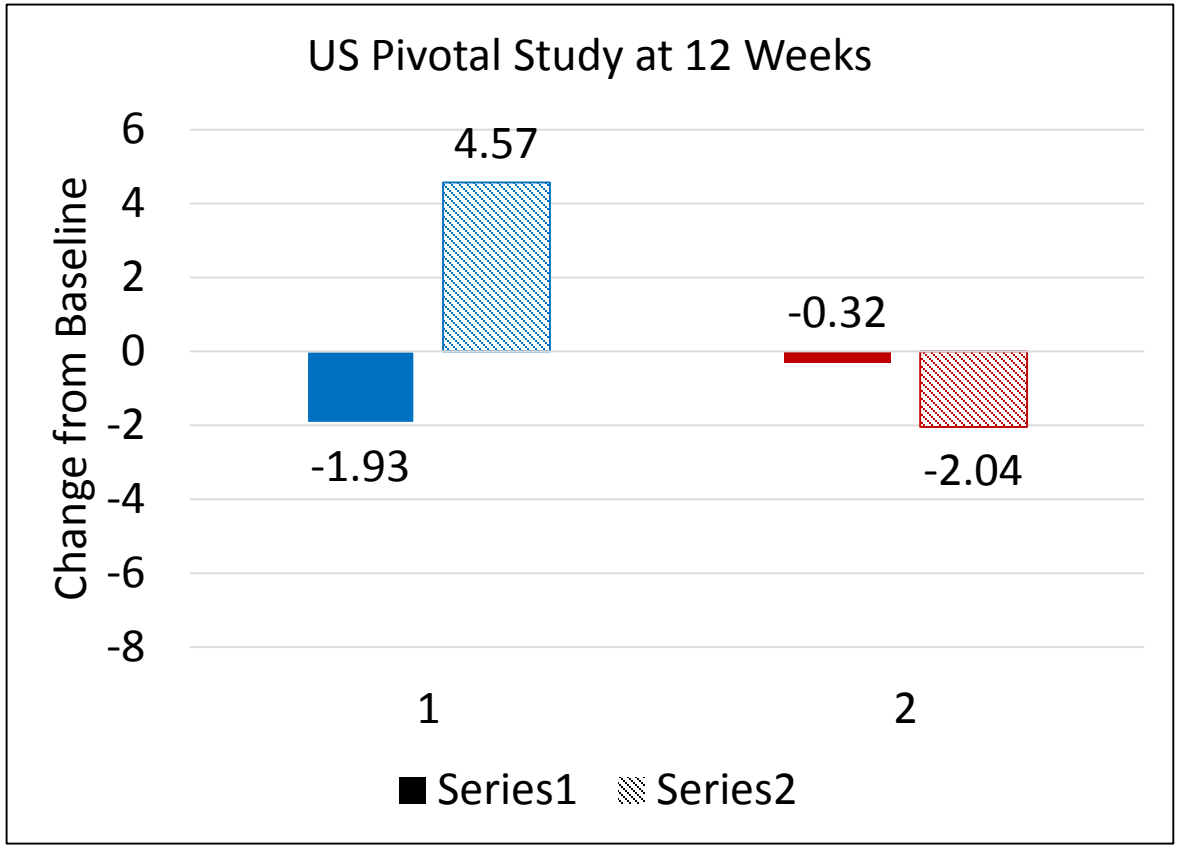
- Presented on a computer touch screen for patient interaction
- Dedicated tasks developed to activate the corresponding brain regions being stimulated by the TMS
- Training difficulty progresses individually based on performance via algorithm

Nonclinical Testing

- The sponsor conducted appropriate non-clinical testing of the system
- All questions regarding non-clinical testing have been resolved.

Performance Between Korea-1 and Pivotal Study Subgroup (Baseline ADAS-Cog ≤30) at 12 Weeks





Neither active nor sham subjects with baseline ADAS-Cog >30 perform as well as those ≤30

Active subjects with baseline ADAS-Cog >30 do not perform as well as those ≤30; however, the sham subjects perform <u>better</u> with a baseline >30