ADMINISTRATIVE INFORMATION

Title: Statistical Analysis Plan for Treating Hyperexcitability in AD with Levetiracetam (LeAD): Neuropsychological Outcome Data

Trial Registration Number: NCT03875638

SAP version: 1.0 March 9, 2026

SAP revisions: none

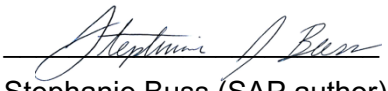
Protocol Version: This SAP corresponds to study protocol dated 4/2/2025.

First Authors: Stephanie Buss, Brice Passera, Recep Ozdemir

Senior Authors: Dan Press, Mo Shafi

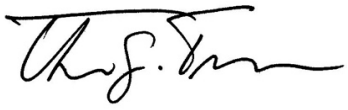
Statistician: Tom Travison

Signatures and dates:



__03/12/2026___

Stephanie Buss (SAP author) date



03/12/2026

Tom Travison (Biostatistician) date

__3/10/2026_____

Mo Shafi (PI) date

INTRODUCTION

Background and rationale:

AD is the most common cause of dementia, affecting 5.5 million people in the US[1]. While amyloid removal therapies are now available, they only slow clinical decline by approximately 30 percent, necessitating identification of new targets for future multimodal interventions. Human studies using Transcranial Magnetic Stimulation (TMS) have demonstrated that patients with AD have cortical hyperexcitability, with decreases in TMS motor threshold[2,3] and abnormal TMS-evoked EEG activity[4,5]. On routine EEG, epileptiform abnormalities are common; epileptiform discharges are seen in up to 42% of patients[6,7], and seizures occur in up to 16% of patients with mild AD in prospective cohort studies[8]. Seizures have been detected using implanted electrodes in AD patients even when no seizures are apparent on scalp EEG[9], suggesting that the true incidence of epilepsy may be even higher. Furthermore, patients with epileptiform discharges or seizures have a faster rate of cognitive decline than those without[6,10], suggesting a fundamental role in disease progression.

Preliminary studies testing levetiracetam (LEV)[11], an antiepileptic drug, in patients with amnestic Mild Cognitive Impairment (aMCI) found improvements in a cognitive task utilizing hippocampal-dependent memory processes, and normalization of BOLD-MRI hippocampal activity[12,13]. In a preliminary study, our group also found that a single dose of intravenous LEV normalizes EEG abnormalities in patients with mild AD[14]. However, it is still unclear whether LEV can normalize hyperexcitability in early AD, at what dose, and if normalization of cortical excitability correlates with improvements in brain function and cognition. In this proposal, using a randomized placebo-controlled crossover design, we will systematically assess if 4 weeks of low dose (125 mg twice daily) versus high-dose (500 mg twice daily) LEV normalizes abnormalities in EEG, TMS and MRI measures of brain function, and improves cognitive function on the Neuropsychological Test Battery (NTB) in participants with either amyloid-positive MCI or mild AD (early AD).

Objectives:

Research hypothesis

The null hypothesis is that there is no difference in cortical hyperexcitability measures or cognitive outcomes between the three treatment conditions (placebo, low-dose LEV, or high-dose LEV). The alternative hypothesis is that there is a difference in cortical hyperexcitability or cognitive outcomes between the three treatment conditions.

Study Objectives

The primary objective of this trial is to determine the extent to which 4-weeks of LEV treatment might improve cognitive function (Objective 1; outlined in this SAP) or normalize cortical hyperexcitability (Objective 2) in early-AD participants.

Secondary objectives are to test whether the beneficial effects of LEV on cognitive function are related to (1) baseline hyperexcitability as indexed by TMS-EEG or MRI measures, and (2) change in cortical excitability and brain function, or (3) the final level of cortical excitability.

Exploratory objectives are to evaluate whether the effects of LEV are related to the presence of baseline epileptiform abnormalities on 24-hour ambulatory EEG or ApoE4 status.

STUDY METHODS

Trial Design:

This trial is a single-center, double-blind, randomized, placebo-controlled crossover design. Participants are randomized in a 1:1:1:1:1:1 ratio to each of three intervention sequences (ABC, ACB, BAC, BCA, CAB, CBA), where A, B, and C represent low dose (125 mg twice daily) versus high-dose (500 mg twice daily) LEV, vs. placebo.

Randomization:

Randomization was completed by the BIDMC Pharmacy and the specific randomization process is contained in secure BIDMC Pharmacy records.

Sample Size:

Enrollment in the study was initially planned to be at least 50 AD participants, with anticipated study completion by 45 participants; however, this sample size was not attained due to the COVID-19 pandemic. Therefore, we report the power attained with the attained sample size (n=23 full completers). This shows the range of effect sizes for which the study is adequately

powered and identifies the minimum detectable effect, providing context for the interpretation of both significant and non-significant results.

Power was computed using G*Power 3.1[15] for a repeated measures ANOVA with within-subject factors as an approximation to the primary linear mixed model with parameters: 3 measurements (treatment conditions), 1 group, n=23 (conservatively, participants completing all 3 treatment periods), α = 0.05, correlation among repeated measures = 0.50, and nonsphericity correction ε = 1. Cohen's *f* was used as the effect size metric.

Effect size benchmarks were derived from meta-analyses of approved AD pharmacotherapies and from prior levetiracetam trials. Table 1 presents the power to detect each benchmark effect size, along with the source literature. Effect size estimates from prior small LEV trials, particularly subgroup analyses, should be interpreted with caution as they are likely inflated due to the winner's curse[16].

Table 1. Statistical Power

Cohen's f	Power	Drug / Study	Effect and Population	Source
0.14	28%	Donepezil	Pooled SMD on ADAS-Cog, AD	Li et al., 2019[17]
0.195	50%	LEV (pooled)	Executive function; RCTs pooled across amnestic MCI, AD, and schizophrenia	Lin et al., 2024[18]
0.24	68%	Galantamine	Pooled SMD on ADAS-Cog, AD	Li et al., 2019[17]
0.28	**82%**	Convention	Power of >80%	
0.83*	>99%	LEV (Vossel, EPI+)	Stroop color-word score in AD with epileptiform activity (n = 9)	Vossel et al., 2021[19]

*G*Power parameters: repeated measures ANOVA, within factors; 3 measurements, 1 group, N = 23, α = 0.05, correlation among repeated measures = 0.50, nonsphericity correction ε = 1. *Vossel et al. reported Cohen's f = 0.83 for the Stroop interference naming subscale in the epileptiform subgroup (n = 9); this is likely inflated due to the small subgroup.*

This study has adequate power (>80%) to detect a small-to-medium effect size (Cohen's *f* = 0.28) and is well-powered (≥90%) for medium effects. For pairwise comparisons between any two treatment conditions, this corresponds to a minimum detectable within-participant difference of approximately 0.61 NTB Composite z-score units (paired t-test approximation, two-tailed). Because the primary LMM additionally adjusts for period-specific pre-treatment scores, the effective sensitivity may be somewhat better than this approximation. However, this study is underpowered to detect effects in the range of approved AD pharmacotherapies (*f* = 0.12–0.14), which is expected for a Phase 2 proof-of-concept trial of this size. A non-significant primary result should be interpreted as inconclusive rather than as evidence of no treatment effect; the estimated effect size and confidence interval will be reported to inform future trial design.

Framework:

The LeAD trial protocol's objective is to test for the superiority of LEV treatment (and dose-responsiveness) compared to placebo on cognitive (objective 1) and neurophysiologic (objective 2) outcome measures in AD during short-term treatment (4 weeks).

Statistical Interim analyses and stopping guidance:

No formal interim analyses for efficacy were planned or performed. The Data Safety Monitoring Board (DSMB) reviewed deidentified adverse event data every 6 months during the trial. The DSMB was blinded to treatment allocation. The DSMB had the authority to recommend early termination if a significant safety signal was identified, based on clinical judgement of the adverse event data. No formal statistical stopping rules were predefined.

Timing of Outcome Assessments: The schedule of procedures is shown in Table 2.

Table 2: Study Schedule for MCI/AD Participants Only

Screening and Baseline					Intervention Period										
	Screen	BL 1	BL 2/ Pre Rx		Period 1 4 Weeks	Post Rx		Pre Rx	Period 2 4 Weeks	Post Rx		Pre Rx	Period 3 4 Weeks	Post Rx	Follow Up
	1 visit	1 v	2 v			2 v		2 v		2 v		2 v		2 v	1 v
Confirm AD dx	X														
CDR	X														
CGIC	X					X		X		X		X		X	
NPI- Q	X				X	X		X	X	X		X	X	X	
C-SSRS	X				X	X		X	X	X		X	X	X	
MMSE	X														
GDS	X														
***Amyloid PET/ Tau PET**	X														
TMS/MRI Screen	X														
Labs	X														
Neuro Exam	X														
Medical Hx & Meds	X														
Blood/Saliva/DNA		X													
Inclusion/Exclusion Review	X	X	X												
****24 hr EEG**		X													
fMRI			X			X		X		X		X		X	
Cognitive Testing			X			X		X		X		X		X	
TMS-EMG-EEG			X			X		X		X		X		X	
Medication Disp.			X					X				X			
Medication Reconciliation						X				X					X
Weekly Phone Call					X				X				X		
****Peak/Trough LEV**			X			X				X				X	
Adverse Events	X	X	X		X	X		X	X	X		X	X	X	

Randomization

4 Week Washout

4 Week Washout

*Amyloid will be measured via PET imaging if not available **If seizure activity noted, the participant will be excluded and referred to their physician for treatment. ***Peak/Trough LEV will be checked during the last week of medication

STATISTICAL PRINCIPLES:

Confidence intervals and p-values:

For all applicable statistical tests, 2-sided tests and an alpha=0.05 will be used. All reported 95% confidence intervals will be 2-sided. Uncorrected p-values will be reported.

Adherence:

Compliance to the randomized treatment during each study period is assessed based on report by the participant and study partner, paired with pill counts and blood tests for LEV levels at the end of each treatment period. Non-compliance during the study period will be defined as missing 9 or more days of treatment (approx. 33% or more) based on participant/study partner report or via end-of-treatment pill count. Descriptive statistics on the percent compliance (N, percent) and Peak and Trough blood LEV levels (mean, SD, minimum, maximum) will be summarized by treatment condition. For any LEV levels below the limit of quantification of our test, the level will be set at half-way between 0 and the limit of detection of the test. Of note, given the limits of quantification of most lab tests, we anticipate that even participants who are compliant with low-dose LEV at 125 mg BID may not have a quantifiable trough level on end-of-treatment blood test.

Protocol Deviations: The following are observed protocol deviations with a direct bearing on the primary outcome:

1) Study medication was held and then restarted in the middle of the study period due to feeling unwell so they could rest and meet with their PCP; post-treatment data was collected after the participant had been restarted on the study medication for the planned 4 weeks (minor)
2) Study medication was inadvertently forgotten for the first 2 weeks of a new treatment period; post-treatment data was collected after the participant had been restarted on the study medication for the planned 4 weeks (minor)
3) Inclusion of a HC participant with an MMSE outside of the range, 27 vs 28; the protocol was later amended to allow inclusion of cognitively normal HC participants with an MMSE of 27 or greater (minor)
4) Blood draw was repeated since creatinine was accidentally omitted from the first blood draw (minor)
5) Participant forgot to hold study medication on the AM of LEV peak and trough, IRB approved a deviation to allow the participant to take an extra day of medication to obtain an accurate peak and trough; the protocol was later amended to allow for greater flexibility in scheduling of the peak and trough dose (minor)
6) An error was made in administration of the MMSE test (an inclusion criterion), an IRB amendment was submitted to allow for a different version of the MMSE to be repeated on a different day to assess for eligibility (minor)
7) A participant was inadvertently included with an MMSE of 18; as CDR was later found to be 0.5, the participant was allowed to continue the study (minor)
8) Participant experienced a seizure during TMS; this participant was allowed to continue the study but did not undergo any additional TMS assessments (MRI and neuropsych assessments only) (minor)
9) Participant experienced discomfort during an MRI scan. For subsequent MRI scans, the participant underwent an abbreviated 30-minute protocol to minimize discomfort but still collect scientific data (minor)
10) Peak/trough LEV values were not collected for a variety of reasons (minor)

All protocol deviations were classified before treatment unblinding. A summary table will report the count and percent of participants with major and minor deviations within each treatment group, along with descriptions of each deviation.

Analysis populations:

The primary analysis will be a modified ITT population consisting of participants who were randomly assigned to treatment and have both a baseline and at least 1 post-baseline measurement of the primary efficacy variable.

A per-protocol sensitivity analysis will be conducted removing any participants who were found to be non-adherent to treatment or if a major protocol deviation occurred to assess robustness of findings.

TRIAL POPULATION:

Recruitment and Eligibility:

A CONSORT diagram will report the number of participants screened, consented, randomized, completed baseline assessments, completed at least 1 treatment period, and completed all 3 treatment periods. Any participants who withdrew or were found to be ineligible will be reported at each study stage.

Withdrawal/follow up:

We will report the number of participants who withdrew 1) within each treatment arm, and 2) after each treatment arm, with a summary of reasons for withdrawal (i.e., side effect of study medication, vs. time-commitment of study, vs. other /life circumstance).

Baseline characteristics:

AD participants' baseline demographics and clinical characteristics will be reported including age, gender, race, ethnicity, years of education, number of APOE4 alleles, AD medications (cholinesterase inhibitors and/or memantine), CDR global, CDR sum-of-boxes, GDS, MMSE, MoCA, WRAT4 Math Computation (as a proxy for academic achievement), cortical atrophy on baseline MRI. Baseline neurophysiologic measures will be reported including rMT, I-O curve inflection point, TMS-EEG parietal excitability, and the presence of epileptiform discharges on 24-hr EEG. Since LEV is available clinically, we will also report the proportion of participants who opted to take LEV after discussion with their physician after the study concluded. Categorical data will be summarized using # and %. Continuous data will be summarized using mean and SD. No statistical analysis is planned for baseline measures, as these are intended as descriptive statistics to describe the participant group.

Compliance and PK Data:

Compliance and pharmacokinetic data will be summarized in a table by treatment condition.

ANALYSIS:

Outcome definitions:

An adapted NTB Composite z-score was prespecified as our study's primary endpoint. Since the study's start, additional data has been published suggesting a greater effect of LEV on fronto-executive function compared to memory in AD, and strongest in patients with greater cortical excitability (based on epileptiform discharges on scalp EEG[19]). Therefore, we expect that if we see a beneficial effect of LEV on cognition, we will see the strongest effect on executive function (Secondary outcome of NTB Executive Function Composite z-score, other

outcomes including Trails B), an intermediate effect with global cognition (Primary outcome of NTB Composite z-score, other outcomes including ADAS-Cog, MoCA), and the smallest effect on memory measures (Secondary outcome of NTB Memory Composite z-score, other outcomes such as RAVLT Delayed Recognition). Furthermore, we expect that any beneficial effects of LEV on cognition will be strongest in AD participants with the greatest cortical excitability. However, it is possible that unexpected patterns in our data could emerge, and therefore we will report all computed results and any observed patterns, with the following pre-specified hierarchy:

Primary Outcome: Neuropsychological Test Battery (NTB) Composite z-score. Participants in this study completed an adapted version of the NTB[20] using 7 tests: Rey Auditory Verbal Learning Test (RAVLT) Immediate Recall, RAVLT Delayed Recall, The Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV) Digit Span, The Controlled Oral Word Association Test (COWAT; FAS), Category Fluency (Animal fluency), Wechsler Memory Scale-Fourth Edition (WMS-IV) Visual Reproduction Immediate Recall, and WMS-IV Visual Reproduction Delayed Recall. Z-scores will be calculated for each of the 7 component tests and averaged together to create an NTB composite z-score. A minimum of 5 of 7 subtests will be required to calculate a valid composite score for each participant at each time point.

Co-Secondary Outcomes: Our hypothesis is that any beneficial effect of LEV will show the greatest effect size in the NTB Executive Function Composite z-score, compared to the NTB Memory Composite z-score.

- **NTB Executive Function Composite z-score:** Calculated by averaging the z-scores of WAIS-IV Digit Span Total Score, COWAT, and Category Fluency for participants with no missing data on at least 2 out of the 3 included tests.
- **NTB Memory Composite z-score:** Calculated by averaging the z-scores of RAVLT Immediate Recall, RAVLT Delayed Recall, WMS-IV Visual Reproduction Immediate Recall, and WMS-IV Visual Reproduction Delayed Recall for participants with no missing data on at least 3 out of the 4 included tests.

Other Cognitive Outcomes: Prior to analysis, all tests will be standardized such that higher scores represent better performance (i.e., some scores such as the ADAS-Cog will be inverted). These analyses are hypothesis-generating; no corrections for multiple comparisons will be applied. Given the number of tests (n=31), we expect that 1-2 tests will show a significant result by chance alone; therefore, we will focus on reporting the effect sizes for the effects of LEV on each test and highlight any patterns that emerge that can help inform future studies, with the following expected interpretations:

- **Global Cognition:** Effects of LEV on these tests will be interpreted as LEV improving global cognition
  - Alzheimer's Disease Assessment Scale—Cognitive Subscale (ADAS-Cog-11 total score)
  - Montreal Cognitive Assessment (MoCA)
- **Daily Function via Caregiver Report:** Effects of LEV on these tests will be interpreted as LEV improving daily functioning as observed by the caregiver/study partner.
  - Clinical Global Impression of Change
  - Functional Activities Questionnaire (FAQ)
  - Alzheimer's Disease Cooperative Study–Activities of Daily Living Scale (ADCS-ADL)
- **Memory – Learning:** Effects of LEV on these tests will be interpreted as LEV improving the ability to encode information during learning tasks.
  - ADAS-Cog Immediate Recall

    - RAVLT Best Learning Trial
    - RAVLT Short Delay Recall
    - RAVLT Immediate Recall
    - WMS-IV Visual Reproduction Immediate Recall (with a visuospatial component)
- **Memory – Recall and Recognition:** Effects of LEV on these tests will be interpreted as LEV improving the ability to store and retrieve information previously learned.
    - ADAS-Cog Delayed Recall score
    - ADAS-Cog Delayed Recognition score
    - RAVLT Long Delay Recall (/15)
    - RAVLT Delayed Recognition (/30, True Positives + True Negatives)
    - RAVLT Percent Retention (Delayed Recall divided by Best Learning Trial)
    - RAVLT Delayed Recall
    - WMS-IV Visual Reproduction Delayed Recall (with a visuospatial component)
    - WMS-IV Visual Reproduction Delayed Recognition (with a visuospatial component)
- **Executive Function – Attention:** Effects of LEV on these tests will be interpreted as LEV improving simple attention.
    - Trails A time
    - WAIS-IV Digit Span: Forward Total
- **Executive Function – Complex:** Effects of LEV on these tests will be interpreted as LEV improving complex executive functions, as outlined below.
    - Stroop Color-Word Raw Score (inhibition)
    - Trails B time (set shifting)
    - Trails B-A time (executive function corrected for working memory)
    - WAIS-IV Digit Span: Backward Total (working memory)
    - WAIS-IV Digit Span Sequencing Total (working memory)
    - ADAS-Cog Maze 3 time (problem solving, with a visuospatial component)
    - ADAS-Cog Maze 4 time (problem solving, with a visuospatial component)
- **Verbal fluency and naming:** Effects of LEV on these tests will be interpreted as LEV improving naming and verbal fluency. If the COWAT shows a significant effect, we will interpret this reflecting improved frontal-lobe mediated word retrieval. If animal fluency shows a significant effect, we will interpret this as reflecting improved medial temporal lobe-mediated word retrieval.
    - COWAT (sum of F, A, and S scores)
    - Category Fluency Test (animals)
    - National Alzheimer's Coordinating Center Uniform Data Set Multilingual Naming Test (MiNT)
- **Tests of Parietal Function:** Effects of LEV on these tests will be interpreted as LEV improving tasks related to both visuospatial function and parietal lobe function. While there is a strong influence of educational attainment on the Wide Range Achievement Test 4 (WRAT4), computational skills can also be sensitive to parietal dysfunction.
    - Benton Judgment of Line Orientation Test (visuospatial function)
    - Tests mentioned above of the ADAS-Cog Maze 3 time, ADAS-Cog Maze 4 time, and Visual Reproduction tests will also be interpreted as containing a visuospatial component and partially reflecting parietal lobe function in addition to executive function, learning, and memory.
    - WRAT4 Math Computation (arithmetic)

NTB z-score calculations: Following the original NTB z-score description[20] and a common approach in the literature[21], we will plan to compute z-scores from the Baseline scores from the AD participants in our study. This will normalize the starting z-scores to 0, so effect sizes in our main model can be easily interpreted as change relative to Baseline.

- Before z-score calculation, data quality checks will be run to assess for approximate normality based on Shapiro-Wilk test and checked for outliers (defined as observations exceeding 3 SDs from the subtest mean). Verified outliers will be Winsorized to the 3 SD boundary prior to calculating normative means and SDs. The proportion of participants at the minimum or maximum score will be calculated for each subtest. If >20% of participants in the normalization sample are at floor or ceiling on any subtest, the SD will be unreliable for that subtest. In such cases, the affected subtest will be excluded from the composite. Additionally, the data will be visualized to see if our modified tele-neuropsychology vs fully in person assessments were systematically different.
- If the quality checks above reveal that the AD baseline data are unsuitable for normalization (i.e., multiple subtests with >10% outliers after data cleaning, non-normality that is not remediable by Winsorization, floor/ceiling effects affecting 2 or more subtests, or large effects of modified tele-neuropsychology testing), the primary analysis will use means and SD from 20 HC recruited under the same study protocol instead. If the HC data are similarly compromised, age-matched published norms will be used (e.g., using the Mayo Older Americans Normative Studies: MOANS published norms, and appropriate manualized normative data). All z-score diagnostics, calculations, and decisions will be made prior to unblinding.

Side Effects and Adverse Events:

Side effects and adverse events were recorded and reported to the DSMB throughout the study in order to allow the DSMB to determine if the trial could ethically be continued as planned. For the NPI-Q and C-SSRS, Peak Week (ie, week with the highest scores) and Final Week scores will be computed and reported for each treatment period. All adverse events will be reported by treatment period.

- Neuropsychiatric Inventory Questionnaire (NPI-Q)
- Columbia Suicide Severity Rating Scale (C-SSRS)
- Adverse events during each study period

Prespecified Subgroup Analyses:

Exploratory analyses will assess the extent to which the efficacy of LEV varies based on:

- Disease stage (MCI vs mild dementia): Prior work using fMRI has suggested a possible inverted U-shaped curve[22], with the greatest hyperexcitability in early-stage MCI compared to more advanced disease stages. However, TMS-EMG findings suggest progressively greater motor cortex hyperexcitability with more advanced disease progression[23,24]. Finding a difference of LEV effectiveness in MCI vs mild dementia subgroups would provide valuable data in designing future phase 3 trials of LEV in AD to target a clinical patient population most likely to benefit.
- APOE4 allele number: The APOE4 allele is thought to affect brain excitability, with the E4 allele associated with increased excitability via loss of GABA-ergic inhibition in mouse models[25]. However, in a prior study of LEV in AD, only the AD subgroup without E4 alleles showed a possible beneficial effect of LEV[26]. Finding a difference in LEV effectiveness based on E4 allele status would advance our understanding of the mechanisms leading to hyperexcitability in AD and guide future phase 3 clinical trials.
- Tele-neuropsychology assessment: A subset of cognitive assessments were administered via modified tele-neuropsychology during the COVID-19 pandemic (visits occurred in person with the same pen and paper tests, but the RA only came into the room intermittently and was otherwise available on 2-way video set up). Tele-

neuropsychological assessment may introduce systematic variance through differences in stimulus presentation, environmental control, and examiner-patient rapport, and the normative data for most cognitive instruments were developed under in-person conditions[27]. Because the proportion of tele-administered assessments in this trial is a direct artifact of COVID-era protocol modifications rather than a participant characteristic, a sensitivity analysis excluding or adjusting for tele-neuropsych visits will evaluate whether testing modality confounds treatment effect estimates.

- Amyloid removal therapies: Given the strong link between amyloid and cortical excitability[28], we will look at subgroups of participants who previously underwent amyloid removal therapy vs. those that did not.
- Baseline excitability: Prior work has suggested that LEV may have the greatest cognitive benefits in executive function in participants with higher excitability, as captured by epileptiform discharges on EEG[19].
    - 24-hour EEG: This is a well-established measure of cortical excitability in clinical care and in research.
    - TMS-EMG Input-Output Curve Inflection Point: This is a relatively novel measure of hyperexcitability that shows increased excitability between AD and age-matched older adults (submitted manuscript under review [29]).
    - TMS-EEG Local Parietal Excitability using source-space modeling: This is a relatively novel measure of hyperexcitability that shows increased excitability in AD in our baseline measures of the present study[30].

Analysis methods:

For each primary and secondary endpoint, separate **linear mixed-effects models (LMM)** with restricted maximum likelihood estimation (REML) will be used to assess the effect of LEV on post-treatment z-score, with fixed effects of pre-treatment z-score, treatment (3 levels), period (3 levels), and random participant-level effects:

$$Post\_Score \sim Pre\_Score + Treatment + Period + (1 \mid Subject\_ID)$$

- **Pre_Score** (continuous): Period-specific pre-treatment z-score. This serves as the ANCOVA-style covariate, adjusting for each participant's cognitive starting point at the beginning of each treatment period. This is a time-varying covariate that changes across periods, not a single enrollment baseline.
- **Treatment** (nominal, 3 levels: Placebo, Low, High): The primary effect of interest. Tests whether cognitive performance differs across the three treatment conditions.
- **Period** (nominal, 3 levels: 1, 2, 3): Accounts for temporal effects including practice effects on repeated cognitive testing, disease progression over the study duration, and other time-related confounds.
- **Subject_ID** (random intercept): Accounts for within-subject correlation arising from repeated measurements on the same participant. Each participant's overall cognitive level is modeled as a random deviation from the population mean. The linear mixed model will be fit using long-format data (one record per participant per completed period) to ensure that participants with incomplete data contribute all available observations to model estimation. After model fitting, the number of included observations and participants will be verified against the enrolled sample to confirm that no data were inadvertently excluded.

Carryover effects: A four-week washout period separates each treatment period. Given the pharmacokinetic properties of LEV (with a plasma half-life of approximately 10-11 hours in older adults[31]), carryover effects are not expected and are not modeled explicitly.

Exploratory cognitive outcomes: The same LMM specification used for the primary outcome will be applied separately to each of the exploratory cognitive outcomes listed, substituting the relevant outcome measure and its period-specific pre-treatment score. Results across all exploratory cognitive outcomes will be summarized in a table reporting the treatment effect estimate (high-dose vs. placebo and low-dose vs. placebo), 95% confidence interval, and uncorrected p-value for each outcome. A forest plot displaying the pairwise treatment effects with 95% confidence intervals across all exploratory outcomes will also be presented to provide a visual summary of the pattern of effects. These analyses are exploratory and hypothesis-generating; no corrections for multiple comparisons will be applied. Significant results should be interpreted with caution given the number of tests conducted.

Safety outcomes: For the NPI-Q and C-SSRS, descriptive statistics (means, standard deviations, medians, ranges) will be reported by treatment condition. If sample size and data distributions permit, the same LMM framework used for efficacy outcomes will be applied, with the omnibus F-test for Treatment evaluated at $\alpha = 0.05$ uncorrected. Any signals will be interpreted as warranting further investigation rather than as definitive conclusions. Adverse events during each study period will be summarized descriptively by treatment condition. The number and percentage of participants experiencing each adverse event type will be tabulated for placebo, low-dose LEV, and high-dose LEV periods.

Primary LMM: Effect of LEV

The primary statistical test is the omnibus F-test for the Treatment term. This 2-degree-of-freedom test evaluates whether there is any effect of treatment condition on post-treatment cognitive performance, without assuming a specific shape for the dose-response relationship. Significance will be evaluated at $\alpha = 0.05$. To facilitate comparison with prior LEV trials and with the power analysis, the omnibus effect size will be reported as Cohen's f.

If the omnibus F-test for Treatment is significant, the pattern of treatment effects will be characterized using pairwise comparisons of least-squares means. These are descriptive follow-ups intended to characterize which treatment conditions drive the overall effect, rather than independent inferential tests:

- Low-dose LEV (125 mg) vs. placebo
- High-dose LEV (500 mg) vs. placebo

For each comparison, the least-squares mean difference, 95% confidence interval, and p-value will be reported. Because these comparisons are gated behind a significant omnibus test and are intended as descriptive characterization, uncorrected p-values will be reported. For each pairwise comparison, a standardized effect size (Cohen's d) will be reported along with the raw mean difference in z-scores to facilitate comparison with published benchmarks for approved AD pharmacotherapies and prior LEV trials.

The pattern of pairwise results will inform interpretation of results:

- Primary hypothesis: If only the low dose differs from placebo, or if both doses improve performance but the low dose shows a larger effect, this would be consistent with the inverted-U hypothesis, in which side effects of LEV at higher doses outweigh the cognitive benefit.

- If both doses improve performance relative to placebo with no significant difference between them, this would suggest a ceiling effect in which even the lower dose is sufficient to normalize hyperexcitability, or in which the cognitive side effects from the higher dose neutralize the benefit obtained from further normalizing hyperexcitability.
- If only the high dose differs from placebo, this would suggest a conventional dose-response relationship in which greater SV2A occupancy produces greater clinical effect.

Sensitivity Analysis: Sequence effects

As a sensitivity analysis, the primary model will be refit including the randomized treatment sequence as a fixed effect (nominal, 6 levels). With six sequence groups and approximately 30 participants, some sequences may contain as few as 3–5 participants, which risks estimation instability or model non-convergence. The primary model omits the Sequence term for this reason, and due to the low likelihood of carryover effects based on LEV half-life.

Post_Score ~ Pre_Score + Treatment + Period + Sequence + (1 | Subject_ID)

If the sequence model converges successfully, the treatment effect estimates will be compared between models. If estimates change by less than 15%, this will be interpreted as confirming that treatment order does not confound the primary results. Both model results will be reported regardless of outcome.

Sensitivity Analysis: LEV PK and exposure responsiveness

To assess whether cognitive outcomes are associated with individual level of drug exposure, secondary analyses will be conducted using data from active treatment periods only. Two measures of drug exposure will be examined: trough LEV blood levels (reflecting sustained steady-state exposure during the treatment period) and peak LEV blood levels (reflecting maximum receptor occupancy). Of note, given the limits of quantification of standard laboratory assays, we anticipate that some participants who are compliant with low-dose LEV at 125 mg BID may not have a quantifiable trough level on end-of-treatment blood draw. For any LEV levels below the limit of quantification, the level will be set at halfway between zero and the limit of detection. If a substantial proportion of trough values fall below the limit of quantification, the trough analysis will be interpreted with appropriate caution. Separate linear mixed models (for Trough and Peak) will test whether greater drug exposure is associated with cognitive benefit:

Post_Score ~ Pre_Score + LEV_level + Period + (1 | Subject_ID)

The slope of the LEV_level term, its 95% confidence interval, and p-value will be reported. A positive slope indicates that higher drug exposure is associated with greater cognitive benefit; a negative slope would suggest diminishing or adverse effects at higher exposure levels.

Visualization: To characterize the shape of the exposure-response relationship without imposing a parametric form, within-period cognitive change (post-treatment minus pre-treatment z-score) will be plotted against end-of-period LEV blood levels for each active treatment period. Individual participant-period observations will be displayed as data points, with a locally weighted smoother (LOWESS) overlaid to aid visual interpretation of the exposure-response curve. These plots will be generated separately for peak and trough LEV levels. If the LOWESS curve suggests a non-linear relationship (e.g., an inverted-U pattern consistent with diminishing benefit at higher exposure), this will be noted and discussed in the context of the categorical dose-response results from the primary analysis.

Sensitivity Analyses: Complete-case sensitivity analysis

To assess whether results are influenced by participants with incomplete data, the primary LMM will be re-fit restricted to participants who completed all three treatment periods.

Nested Covariate Analyses

To examine whether baseline characteristics predict treatment outcome or moderate the treatment effect, separate models will be fit adding each of the following covariates individually. These analyses are exploratory and hypothesis-generating; no corrections for multiple comparisons will be applied.

For each covariate, a nested model comparison approach will be used:

1. **Main effect model:** The covariate is added as a main effect to the primary model. A likelihood ratio test (ML estimation) compares this model to the primary model to assess whether the covariate predicts overall cognitive outcome.
2. **Interaction model:** A Treatment × Covariate interaction term is added. A likelihood ratio test compares this model to the main effect model to assess whether the covariate moderates the treatment effect (i.e., whether treatment response differs by subgroup). This is the primary test of interest for each covariate.

Covariates of interest:

- **Baseline CDR-SB**
- **Baseline MMSE**
- **Baseline GDS**
- **Gender**
- **Age**
- **Education**
- **APOE4 allele number**
- **Baseline I-O Curve Inflection Point**
- **Baseline TMS-EEG Parietal Excitability**
- **Freesurfer MRI hippocampal volume**
- **Freesurfer MRI whole brain cortical thickness**

Note: Each covariate will be tested individually rather than simultaneously.

Exploratory Subgroup Analyses

Treatment effects will be examined within pre-specified subgroups by fitting the primary model separately within each stratum.

Post_Score ~ Pre_Score + Treatment + Period + (1 | Subject_ID)

Subgroup-specific treatment effects (least-squares mean differences with 95% confidence intervals) will be presented in a forest plot alongside the overall treatment effect from the primary analysis. Formal tests of Treatment × Subgroup interactions will also be conducted in the full sample to assess whether treatment effects differ significantly between subgroups.

Pre-specified subgroups:

Cognitive Stage: MCI vs. Mild Dementia (CDR 0.5 vs. 1.0)

Participants will be classified by baseline Clinical Dementia Rating (CDR) score. This subgroup analysis examines whether treatment effects differ by disease stage.

APOE4 Allele Count (0 vs. 1 vs. 2)

Participants will be classified by APOE4 carrier status. If sample sizes permit, a three-way split (0, 1, 2 alleles) will be examined; otherwise, a binary split (non-carrier vs. carrier) will be used.

Tele-neuropsychology (Yes/No)

To assess the impact of testing modality on results, we will examine the effects of LEV on participants with fully in person testing vs. modified tele-neuropsychology testing.

Exposure to amyloid-targeted monoclonal antibodies (Yes/No)

To assess the impact of prior exposure to amyloid removal therapies, we will examine the effects of LEV in participants who had no exposure to amyloid-targeted monoclonal antibodies vs. those who previously underwent therapy with aducanumab, lecanemab, or donanemab.

Epileptiform Discharges on 24-Hour EEG (Yes/No)

Participants will be classified by the presence or absence of epileptiform discharges on 24-hour ambulatory EEG. This subgroup analysis tests the hypothesis that participants with evidence of network hyperexcitability may derive greater benefit from LEV's anti-seizure and synaptic vesicle modulation mechanisms.

TMS Motor Input-Output Curve Inflection Point (Tertiles)

Participants will be classified by baseline cortical excitability as measured by the inflection point of the TMS motor cortex input-output (I-O) curve. Will be divided into tertiles (approximately n = 8 in each group). I-O curve fitting will be performed as described in our submitted publication[29]. Scalp-to-cortex distance (SCD) will be included as a covariate as in our prior analyses. This analysis tests the mechanistic hypothesis that participants with greater baseline cortical excitability are more likely to benefit from LEV.

TMS-EEG Parietal Cortical Excitability (Tertiles)

Participants will be classified by baseline parietal cortical excitability as measured by TMS-EEG using source space modeling. We will use the same ROIs and time windows as used in our publication from this same dataset[30]. This analysis provides a complementary measure of cortical excitability in a disease-relevant brain region.

Interpretation of Tertiles

Interpretation of TMS-EMG and TMS-EEG tertile results will be guided by the following pre-specified framework in order to identify patterns which are biologically plausible. The most plausible outcome is a threshold effect. For example, if the highest-excitability tertile shows a treatment response that is qualitatively distinct from the lower two tertiles, this would suggest that levetiracetam benefits only those patients whose excitability exceeds a pathological threshold. This would be consistent with the hypothesis that the LEV acts by normalizing aberrant hyperexcitability rather than suppressing normal neural activity. Second, a monotonic dose-response pattern, in which treatment benefit increases (or decreases) progressively across tertiles. This would suggest a graded relationship between baseline excitability and LEV

responsiveness and would support the use of excitability as a continuous moderator in future trials. A U-shaped pattern — in which the middle tertile shows a different treatment response than both extremes — would be considered inconsistent with our mechanistic model and likely attributable to sampling variability in a small study.

Exploratory Analyses of Input-Output Curve and TMS-EEG Parietal Cortical Excitability

Since this is among the first investigations of TMS-EMG or TMS-EEG derived excitability as a pharmacodynamic moderator in a levetiracetam trial, additional exploratory analyses may be conducted if the tertile results suggest patterns not fully captured by the pre-specified framework. These may include alternative cutpoints (e.g., median split), examination of the continuous excitability measure as a moderator, or post-hoc visualization of individual treatment responses plotted against baseline excitability. Any such analyses will be clearly labeled as exploratory and hypothesis-generating.

Model Diagnostics:

The following diagnostic checks will be performed on the primary model:

- **Residual normality:** Assessed via histogram and Q-Q plot of residuals. Shapiro-Wilk test will be reported.
- **Homoscedasticity:** Residuals will be plotted against fitted values, treatment condition, and period to assess variance homogeneity.
- **Influential observations:** Cook's distance and leverage values will be examined. Given the small sample size, individual observations may exert undue influence on estimates.
- **Random effects distribution:** The distribution of participant-level random intercepts will be assessed for normality.

If residual diagnostics reveal substantial departures from normality or influential observations, additional sensitivity analyses will be conducted, including refitting the primary model after excluding influential observations identified by residual diagnostics.

Missing data:

The LMM uses all available observations under a missing-at-random (MAR) assumption, estimated via maximum likelihood. Participants who completed fewer than three treatment periods contribute data from their completed periods and retain their originally randomized sequence assignment. No imputation of missing outcomes will be performed.

The pattern and extent of missing data will be described, including the number of participants completing one, two, or three treatment periods, and whether dropout was associated with specific treatment conditions, sequence assignments, or baseline characteristics. If dropout appears non-random with respect to treatment condition, this will be noted as a potential limitation.

Statistical software: JMP version 18 will serve as the primary statistical software. R and Matlab may be used for supplementary analysis if needed.

References:

1. 2017 Alzheimer's disease facts and figures. *Alzheimers Dement. J. Alzheimers Assoc.* **13**, 325–373 (2017).

2. Ferreri, F. *et al.* Motor cortex excitability in Alzheimer's disease: a transcranial magnetic stimulation study. *Ann. Neurol.* **53**, 102–108 (2003).
3. Di Lazzaro, V. *et al.* Motor cortex hyperexcitability to transcranial magnetic stimulation in Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatry* **75**, 555–559 (2004).
4. Julkunen, P. *et al.* Navigated TMS combined with EEG in mild cognitive impairment and Alzheimer's disease: a pilot study. *J. Neurosci. Methods* **172**, 270–276 (2008).
5. Ferreri, F. *et al.* Sensorimotor cortex excitability and connectivity in Alzheimer's disease: A TMS-EEG Co-registration study. *Hum. Brain Mapp.* **37**, 2083–2096 (2016).
6. Vossel, K. A. *et al.* Incidence and impact of subclinical epileptiform activity in Alzheimer's disease. *Ann. Neurol.* **80**, 858–870 (2016).
7. Horváth, A. *et al.* Prevalence, Semiology, and Risk Factors of Epilepsy in Alzheimer's Disease: An Ambulatory EEG Study. *J. Alzheimers Dis. JAD* **63**, 1045–1054 (2018).
8. Romanelli, M. F., Morris, J. C., Ashkin, K. & Coben, L. A. Advanced Alzheimer's disease is a risk factor for late-onset seizures. *Arch. Neurol.* **47**, 847–850 (1990).
9. Lam, A. D. *et al.* Silent hippocampal seizures and spikes identified by foramen ovale electrodes in Alzheimer's disease. *Nat. Med.* **23**, 678–680 (2017).
10. Vossel, K. A. *et al.* Seizures and epileptiform activity in the early stages of Alzheimer disease. *JAMA Neurol.* **70**, 1158–1166 (2013).
11. Sanchez, P. E. *et al.* Levetiracetam suppresses neuronal network dysfunction and reverses synaptic and cognitive deficits in an Alzheimer's disease model. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2895-2903 (2012).
12. Bakker, A. *et al.* Reduction of hippocampal hyperactivity improves cognition in amnestic mild cognitive impairment. *Neuron* **74**, 467–474 (2012).

13. Bakker, A., Albert, M. S., Krauss, G., Speck, C. L. & Gallagher, M. Response of the medial temporal lobe network in amnestic mild cognitive impairment to therapeutic intervention assessed by fMRI and memory task performance. *NeuroImage Clin.* **7**, 688–698 (2015).
14. Musaeus, C. S., Shafi, M. M., Santarnecchi, E., Herman, S. T. & Press, D. Z. Levetiracetam Alters Oscillatory Connectivity in Alzheimer's Disease. *J. Alzheimers Dis. JAD* **58**, 1065–1076 (2017).
15. Faul, F., Erdfelder, E., Buchner, A. & Lang, A.-G. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* **41**, 1149–1160 (2009).
16. Button, K. S. *et al.* Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **14**, 365–376 (2013).
17. Li, D.-D., Zhang, Y.-H., Zhang, W. & Zhao, P. Meta-Analysis of Randomized Controlled Trials on the Efficacy and Safety of Donepezil, Galantamine, Rivastigmine, and Memantine for the Treatment of Alzheimer's Disease. *Front. Neurosci.* **13**, (2019).
18. Lin, C.-Y., Chang, M.-C. & Jhou, H.-J. Effect of Levetiracetam on Cognition: A Systematic Review and Meta-analysis of Double-Blind Randomized Placebo-Controlled Trials. *CNS Drugs* **38**, 1–14 (2024).
19. Vossel, K. *et al.* Effect of Levetiracetam on Cognition in Patients With Alzheimer Disease With and Without Epileptiform Activity: A Randomized Clinical Trial. *JAMA Neurol.* **78**, 1345–1354 (2021).
20. Harrison, J. *et al.* A Neuropsychological Test Battery for Use in Alzheimer Disease Clinical Trials. *Arch. Neurol.* **64**, 1323–1329 (2007).
21. Harrison, J. E. *et al.* Psychometric Evaluation of the Neuropsychological Test Battery in Individuals with Normal Cognition, Mild Cognitive Impairment, or Mild to Moderate Alzheimer's Disease: Results from a Longitudinal Study. *J. Prev. Alzheimers Dis.* **5**, 236–244 (2018).
22. Celone, K. A. *et al.* Alterations in memory networks in mild cognitive impairment and Alzheimer's disease: an independent component analysis. *J. Neurosci. Off. J. Soc. Neurosci.* **26**, 10222–10231 (2006).

23. Chou, Y., Sundman, M., That, V. T., Green, J. & Trapani, C. Cortical Excitability and Plasticity in Alzheimer's Disease and Mild Cognitive Impairment: A Systematic Review and Meta-analysis of Transcranial Magnetic Stimulation Studies. *Ageing Res. Rev.* **79**, 101660 (2022).
24. Zadey, S. *et al.* Higher motor cortical excitability linked to greater cognitive dysfunction in Alzheimer's disease: results from two independent cohorts. *Neurobiol. Aging*.
25. Najm, R., Jones, E. A. & Huang, Y. Apolipoprotein E4, inhibitory network dysfunction, and Alzheimer's disease. *Mol. Neurodegener.* **14**, 24 (2019).
26. The HOPE4MCI study: A randomized double-blind assessment of AGB101 for the treatment of MCI due to AD - Mohs - 2024 - Alzheimer's & Dementia: Translational Research & Clinical Interventions - Wiley Online Library. https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/trc2.12446.
27. Marra, D. E., Hamlet, K. M., Bauer, R. M. & Bowers, D. Validity of Teleneuropsychology for Older Adults in Response to COVID-19: A Systematic and Critical Review. *Clin. Neuropsychol.* **34**, 1411–1452 (2020).
28. Targa Dias Anastacio, H., Matosin, N. & Ooi, L. Neuronal hyperexcitability in Alzheimer's disease: what are the drivers behind this aberrant phenotype? *Transl. Psychiatry* **12**, 257 (2022).
29. Cho, J. *et al.* Input-output curves demonstrate motor cortex hyperexcitability associated with cognitive impairment in Alzheimer's disease.
30. Ozdemir, R. A. *et al.* Neurophysiological signatures of default mode network dysfunction and cognitive decline in Alzheimer's disease. *Sci. Adv.* **11**, eadt8991 (2025).
31. Hernández-Mitre, M. P. *et al.* Population Pharmacokinetics and Dosing Recommendations of Levetiracetam in Adult and Elderly Patients With Epilepsy. *J. Pharm. Sci.* **109**, 2070–2078 (2020).