Statistical Analysis Plan

1. Administrative information

1.1. Title and Trial Registration

Full study title: Optimizing Outcomes of Treatment-Resistant Depression in Older Adults

Acronym: OPTIMUM

ClinicalTrials.gov identifier: NCT02960763

Washington University in St. Louis IRB ID number: 201609085

PCORI Contract number: TRD-1511-33321

Additional funding by NIMH ("OPTIMUM NEURO" collaborative R01)

SAP version: 4.0

Study protocol version: 3.0 (3/4/19)

SAP version history: 1.0 (12/26/19), 2.0(2/3/20), 3.0 (5/26/20), 4.0 (11/2/2021)

SAP version justification: -

SAP Revision timing: Revision will be conducted after reviewing by PI and biostatistician. This version is the primary draft of the detailed SAPs for each outcome.

1.2. Roles and Responsibility

Author: Lei Yang, Statistical Data Analyst

Biostatistician: J. Philip Miller

Principal investigator: Eric Lenze, MD

SAP contributors and roles: Michael Yingling, Data Manager

Vy Pham, Statistical Data Analyst

Emily Lenard, Study Coordinator

Affiliations: 1. Washington University in St. Louis, Missouri, USA

2. University of California Los Angeles, California, USA

3. University of Pittsburgh, Pennsylvania, USA

4. Columbia University, New York, USA

5. University of Toronto, Ontario, Canada

Table of contents

1. Administrative information 1

1.1. Title and Trial Registration 1

1.2. Roles and Responsibility 1

Table of contents 2

Abbreviation 4

2. Introduction 5

2.1. Background and Rationale: 5

2.2. Objectives 5

2.2.1. Objectives and study aims 5

2.2.2. Hypothesis 5

2.2.3. Scope 6

3. Study Methods 6

3.1. General Study Design and Plan: 6

3.2. Randomization 6

3.3. Sample Size 7

3.4. Framework 7

3.5. Statistical interim analyses and stopping guidance 7

3.6. Timing of final analysis 7

3.7. Timing of outcome assessment 7

4. Statistical principles 7

4.1. Confidence intervals and p values 7

4.2. Adherence and protocol deviations 7

4.2.1. Definitions of protocol deviations 7

4.2.2. Adherence and protocol deviations to be summarized 8

4.2.3. Electronic Data Capture 8

4.3. Analysis populations 8

5. Trial Population 8

5.1. Screening data 8

5.2. Eligibility 8

5.2.1. Inclusion Criteria (Steps 1 & 2) 8

5.2.2. Exclusion Criteria (Steps 1 & 2) 9

5.2.3. Exclusions to Enter Step 2 9

5.3. Recruitment 10
5.4. Withdrawal/follow-up 10
5.4.1. General withdrawal rules 10
5.5. Baseline participant characteristics 11
5.5.1. List of baseline characteristics 11
5.5.2. Descriptive summarization of baseline participant characteristics 11
6. Analysis 12
6.1. Outcome definitions 12
6.1.1. Primary and secondary outcomes 12
6.1.2. Outcome measurements 12
6.2. Analysis methods 13
6.3. Missing data 13
6.5. Harms 14
6.6. Statistical software 14

Abbreviation

*ADM	Antidepressant Medication
AE	Adverse Event
ATHF	Antidepressant Treatment History Form
*CFR	Code of Federal Regulations
CIRS-G	Cumulative Illness Rating Scale-Geriatric
DAST-10	Drug Abuse Screening Test
DSMB	Data Safety and Monitoring Board
EMR	Electronic Medical Record
*E-CRF	Electronic Case Report Form
EKG/ECG	Electrocardiogram
*FDA	Food and Drug Administration
GCP	Good Clinical Practice
*HIPAA	Health Insurance Portability and Accountability Act
*ICF	Informed Consent Form
IRB	Institutional Review board
ITT	Intention-to-treat
MADRS	Montgomery-Asberg Depression Rating Scale
MD – (1-5)	Standards for Preventing and Handling Missing Data
MDD	Major Depressive Disorder
MDE	Major Depressive Episode
MINI	Mini-International Neuropsychiatric Interview
*NMS	Neuroleptic Malignant Syndrome
OPTIMUM	Optimizing Outcomes of Treatment-Resistant Depression in Older Adults
PCORI	Patient Centered Outcomes Research Institute
PHQ-9	Patient Healthcare Questionnaire
PI	Principal Investigator
PROMIS	Patient Reported Outcome Measurement Information System
RCT	Randomized Controlled Trial
REDCap	Research Electronic Data Capture
SAB	Stakeholder Advisory Board
SAE	Serious Adverse Event
TRD	Treatment-Resistant Depression

*Appeared in the protocol, but not in the SAP

2. Introduction

2.1. Background and Rationale:

One-half or more of older adults fail to remit with antidepressant pharmacotherapy. Treatment-resistant depression (TRD) in older adults is highly deleterious because persistence of depression is a leading cause of disability, suicide, dementia, and premature mortality. Making it worse is the lack of evidence-supported treatments at a stage in life when medications' benefit vs. risk ratio crucial.

The Optimizing Outcomes of Treatment-Resistant Depression in Older Adults (OPTIMUM) study will provide the evidence that older adults need to get an effective treatment that works best for them, improving their quality of life while minimizing risks of medications.

2.2. Objectives

2.2.1. Objectives and study aims

The study aims to close the evidence gap on late-life depression in three ways. First, we will examine the comparative benefits and risks of antidepressant strategies (augmentation and switching) in older adults with TRD. In addition to looking at rates of depression remission, we will focus on the effects on the outcomes that also matter to older adults, like well-being, adverse events, and falls. Secondly, we will explore how aging changes the balance of benefits vs. risks. With aging comes a decline in brain and systemic health that may alter the benefit/risk ratio of antidepressant strategies. Third, we will maximize stakeholder engagement by conducting semi-structured qualitative interviews of patients and providers and convening a Stakeholder Advisory Board (SAB) to guide the study and ensure its relevance to patients and providers Armed with the knowledge of differential benefits and risks, stakeholders could provide personalized precision care that maximizes the benefits of TRD treatment strategies for older adults with minimizing risks.

2.2.2. Hypothesis

Theme 1: Benefits and risks of antidepressants: outcomes that matter for patients and clinicians

1. Effectiveness. Augmentation arms will show greater improvement than monotherapy arms for effectiveness outcomes.
    a. Wellness outcome: Improved psychological well-being. (H1a)
    b. Symptom outcome: Remission from depression (defined as MADRS $\leq$10 at Step end). (H1b)
2. Safety. Augmentation arms will have more tolerability and safety concerns (i.e., more serious adverse events) than monotherapy arms.
    a. Serious adverse events (H2a), falls, and fall-related injuries (H2b) are the safety outcomes.

The primary analytic strategy is a test of specific treatment strategies (e.g., aripiprazole augmentation). More recently, a report on these strategies in mixed-age veterans with TRD (Mohamed et al, JAMA 2017) has suggested that augmentation approaches yield higher remission rates than does switch. The TRD field has, likewise, converged on the overarching question of whether augmentation is superior to switch, more so than the specific question of which augmentation approach is superior (e.g., Ruberto, Jha, & Murrough, "Pharmacological treatments for patients with treatment-resistant depression", Pharmaceuticals 2020; Voineskos, Daskalakis & Blumberger, "Management of treatment-resistant depression", Neuropsychiatr Dis Treat 2020) with a recent Cochrane systematic review (Davies et al, 2019) determining that augmentation strategies appeared efficacious while switch strategies had uncertain efficacy.

For this reason, a secondary analysis of each hypothesis will test the overall strategy of augmentation vs. switch. The two augmentation arms in Step 1 will be summed and compared to the switch arm.

Theme 2: Aging influences the balance of benefits and risks of antidepressant strategies.

3. Moderator analyses (Heterogeneity of Treatment Effect). Exploratory analysis (HT-1): Effectiveness and safety differences between treatment arms will be moderated by age. Specifically, with increasing age effectiveness differences between treatment arms will decrease, and safety differences between treatment arms will increase.

2.2.3. Scope

This Statistical Analysis Plan will be the guiding document for the quantitative analyses that will be conducted in the OPTIMUM study.

3. Study Methods

3.1. General Study Design and Plan:

Participants will be recruited from five regions with diverse and representative populations encompassing approximately 708 adults. The five regions are: St. Louis and rural Missouri; Los Angeles City and County; Pittsburgh and Western Pennsylvania; New York City; and Toronto and rural Ontario. Participants will be non-demented adults aged 60+, with equal proportions aged 60-70 and 70+, with current major depression that has failed to respond to 2+ adequate antidepressant trials. Two-thirds of the sample will be women and 86% White, 7% Black, 8% Latino/Hispanic, and 4% other racial groups.

All recruited adults will be randomized to 10 weeks of one of three Step 1 strategies: aripiprazole augmentation, bupropion augmentation, or switch to bupropion. Those who do not attain remission in Step 1 or those who do not qualify for Step 1 will be randomized to 10 weeks of one of two Step 2 strategies: lithium augmentation or switch to nortriptyline. Those who complete acute treatment will be followed in a one-year continuation, with assessments every four months. This pragmatic RCT will be carried out in real-world clinical settings. Primary care and mental health clinical partners will provide treatments, with decision support from the study team.

Patients will be identified from clinical networks at each of the five participating centers. Our two primary mechanisms of recruitment will be through screening and referrals. We will screen through clinical networks using Electronic Medical Records (EMR) and will contact all patients in practices aged 60 and older who in the recent past were prescribed or are currently taking antidepressants. Referrals will be elicited from practitioners and through advertisements to patients and their caregivers.

An additional mechanism of recruitment will be through IRB-approved hospital network registries and social media advertisements and posts on Facebook and Instagram. Potential participants who see Facebook ads and want to participate in the study can click on the Facebook ad and it will direct them to a secure and encrypted survey website using Redcap to collect name, phone number, and email address. Research study staff will contact the potential participants to screen them for study eligibility.

3.2. Randomization

Randomization of participants will be implemented in REDCap, using a randomized block design by alternating blocks of 3 and 6 within strata by site (i.e., St Louis vs. Pittsburgh), referral source (primary care vs. specialty mental health) and age (<70 vs >70).

Only assessors of outcome measures will be blinded. Assessors will be instructed to not ask, and participants instructed to not share the treatment to which they were assigned. Participants, providers, and other research staff will not be blinded to treatment conditions.

In this pragmatic design where all levels of adherence are allowed, participants will be considered randomized assuming they meet the eligibility requirements at screening (including any safety screens as applicable) and they have been given a randomization assignment (i.e., even if participants never took a dose of the randomized medication, they will be included in the analysis).

3.3. Sample Size

Our original proposal was to randomize 1,500 patients in Step 1 and 800 for Step 2. This was adjusted mid-study when it became clear that recruitment efforts would not achieve this sample size. Newly adjusted sample size is 708 (estimated mid-study), including 123 participants who were enrolled into step 2 without first participating in step 1.

3.4. Framework

Hypothesis testing framework: H_0 and H_1 with confidence intervals for differences.

3.5. Statistical interim analyses and stopping guidance

No interim analysis is planned. As there is no placebo and all treatments are FDA-approved and frequently used in older adult, no contingency plans for early stopping because of futility or safety are planned.

3.6. Timing of final analysis

Analyses will ensue after recruitment is completed and all participants have finished their acute phase.

This statistical analysis plan was added to the study protocol at clinicaltrials.gov, before the closure of the database and before any analyses had been conducted. Independent study monitoring was conducted in adherence to the Good Clinical Practice(GCP) guidelines. The statistician will be blind to treatment assignments during the analysis.

3.7. Timing of outcome assessment

Outcome assessment will be conducted at the end of the acute phase for both step 1 and step 2. During the continuation phase, assessments will be conducted at month 4, month 8, and month 12.

4. Statistical principles

4.1. Confidence intervals and p values

To test our primary hypothesis (H1a), a significance level of .05/3 will be considered as statistically significant in Step 1. A significance level of .05 will be considered statistically significant in Step 2. If the significance level is greater than .05, we will emphasis the increased chance of a type I error. Results will be presented with their values (e.g. regression coefficients, standard deviation, etc.) with 95% confidence intervals.

4.2. Adherence and protocol deviations

4.2.1. Definitions of protocol deviations

Any alteration or modification to the IRB-approved research without prospective IRB approval. The term research encompasses all IRB-approved materials and documents including the detailed protocol, IRB

application, consent form, recruitment materials, questionnaires/data collection forms, and any other information relating to the research study.

4.2.2. Adherence and protocol deviations to be summarized

As a pragmatic study, all levels of adherence are allowed. Compliance with the study-recommended medication and dosing will be assessed at each decision support call using the Protocol Adherence Form. Reasons for non-compliance will be documented.

4.2.3. Electronic Data Capture

This study will utilize REDCap-based electronic Case Report Forms (e-CRF's) for data capture and storage. Source documents, including laboratory values, medical records and physician communications will be maintained at each site to verify adherence to protocol, inclusion/exclusion criteria, and data accuracy.

4.3. Analysis populations

We will follow the intention-to-treat (ITT) principle. In other words, patients who receive co-prescribed medications, stop randomized medication, or switch to a different medication will continue to be followed. To confirm that ITT findings represent the "true" data on effectiveness, safety, and tolerability of management strategies, we will conduct a sensitivity analysis that will compare ITT to per-protocol findings.[1]

5. Trial Population

5.1. Screening data

Potential participants identified through each site's IRB-approved recruitment plans will be screened for eligibility. Staff and PI's at each site are responsible for verifying that participants at their site meet all eligibility criteria before being randomized to acute treatment. We anticipate differences between sites in which assessments may be obtained in pre-screening (pre-written consent) versus screening (post written consent). Despite these potential differences, the following assessments must be conducted prior to randomization to ensure eligibility (see also Section 5.2.1 and 5.2.2 for Inclusion and Exclusion Criteria):

- PHQ-9
- MINI MDE Module
- Screening questions for treatment resistance and medical exclusions
- Short Blessed Test
- ATHF
- CIRS-G

5.2. Eligibility

All participants will meet the following eligibility criteria:

5.2.1. Inclusion Criteria (Steps 1 & 2)

a) Men and women aged 60 and older.

[1] In the course of conducting analyses of effectiveness, the research team noted that many participants had a prior trial of one of the randomized medications. This was allowed per the study protocol. Because of these, a sensitivity analysis was conducted examining remission rates and change in depression scores for only those participants who had not had a prior trial of one of the randomized treatments. This was only done for Step 1, because Step 2 treatments were rarely used in prior trials.

b) Current Major Depressive Disorder (MDD), single or recurrent, as diagnosed by DSM-5 criteria.

c) Failure to respond adequately to two or more antidepressant treatment trials of recommended dose and length (approximately 12 weeks; minimum 4 weeks at adequate dose).

d) PHQ-9[2] score of 10 or higher.

5.2.2. Exclusion Criteria (Steps 1 & 2)

a) Inability to provide informed consent.

b) Dementia, as defined by Short Blessed ≥10 and/or clinical evidence of dementia. Patients screened out due to possible dementia will be referred to a local Memory Clinic or back to their clinician for evaluation to clarify the presence or absence of dementia.

c) Lifetime diagnosis of bipolar I or II disorder, schizophrenia, schizoaffective disorder, schizophreniform disorder, delusional disorder, or current psychotic symptoms. A recommendation for psychiatric referral will be made in these cases.

d) High risk for suicide (e.g., active suicide ideation and or current/recent intent or plan) and unable to be managed safely in the clinical trial, such as unwilling to be hospitalized). Urgent psychiatric referrals will be made in these cases.

e) Contraindication to proposed study medications, as determined by study physician including history of intolerance or non-response to proposed medications.

f) Non-correctable, clinically significant sensory impairment (e.g., cannot hear well enough to cooperate with interview).

g) Unstable medical illness, including delirium, uncontrolled diabetes mellitus, hypertension, hyperlipidemia, or cerebrovascular or cardiovascular risk factors that are not under medical management. This will be determined based on information from the patient's personal physician's and study physician clinical judgment. Referral to the patient's personal physician or to a general practitioner will be made in these cases.

h) Moderate to severe substance or alcohol use disorder, as determined by study physician. Referral to appropriate treatment will be made in these cases.

i) Seizure disorder.[3]

j) Parkinson's Disease[3]

No exclusion criteria are based on race, ethnicity, or gender.

5.2.3. Exclusions to Enter Step 2[4]

The following conditions are contraindications to Step 2 medications. Participants with them will not be eligible for Step 2 participation (but may be considered for Step 1 provided they meet criteria outlined in study protocol Sections 2.a & 2.b.)

---

[2] Initially the investigators used a score of 6 or greater but, per discussion with PCORI methodology team, changed this to 10 or greater during the study.
[3] When participants were allowed direct entry to Step 2, exclusion criteria i) and j) were exclusion for Step 1 only.
[4] The investigators allowed eligible individuals to directly enter Step 2 (i.e. if ineligible for Step 1). Thus, participants could enter from Step 1 or directly from initial consent into the study.

a) QTC prolongation or Wide QRS on EKG

b) Active Ischemic Heart Disease as evidenced by angina or requiring treatment (e.g., nitrates) for ischemic attacks. Patients with history of prior MI, stent, or bypass may be included who have had no symptoms of ischemia (e.g., no chest pain) for 2 years.

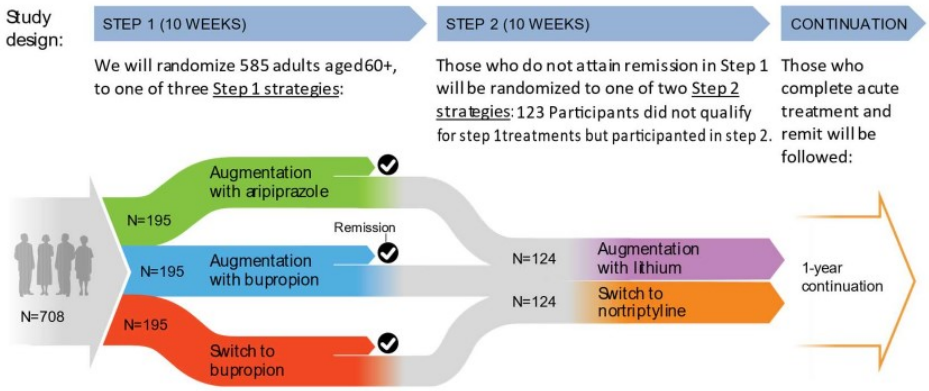
c) Acute or chronic renal insufficiency (as indicated by creatinine clearance below 30 mL/min; suspected if creatinine above 1.5 mg/dL; per PI and/or clinician discretion)

d) Narrow-angle glaucoma

5.3. Recruitment

A Consort diagram will be used to visualize the flow of patients. We will report the number of eligible participants and the number of screened participants. The reasons for lack of eligibility will be enumerated. The number randomized will be reported as well as the reasons for non-randomization. The number randomized into each of the three arms (augmentation with aripiprazole, augmentation with bupropion, and switch to bupropion) will be reported. For those who do not remit under step 1 will be randomized in step 2 to one of 2 arms (augmentation with lithium and switch to nortriptyline). See Figure 1.

Figure 1.



*sample size estimated mid-study

5.4. Withdrawal/follow-up

5.4.1. General withdrawal rules

Participants will be asked to continue in the study for follow-up regardless of their and their physician's adherence to the medication recommendations from the study team. Participants who discontinue study medications for tolerability reasons will be asked to complete the appropriate Acute Phase End Assessments for their phase of the study. Participants who end Step 1 treatment early due to tolerability issues will be invited to proceed to Step 2 after completing Acute Phase End Assessments.

Participants who choose to withdraw from the study early will be asked to complete the appropriate final visit for their phase of the study (Step 1 acute endpoint, Step 2 acute endpoint, or Continuation Month 12).

Withdrawal from the study for any reason, including study completion, will be documented in the Electronic Case Report Form.

Participants' reasons for withdrawal will be summarily reviewed by each site's IRB at Continuing Review and at DSMB meetings.

5.5. Baseline participant characteristics

5.5.1. List of baseline characteristics

Informed consent forms will be signed by participants at baseline. Local research team assessments elements collected during baseline:

- MINI MDD & Bipolar Modules
- PHQ-9 (If more than 7 days has elapsed since the PHQ-9 was done at pre-screen.)
- Suicide History, Risk & Protective Factors
- Baseline Side Effects
- Expectations for Adherence
- Falls & Fall-related Injuries (history)
- Height and Weight (self-reported)
- If indicated by pre-screen: DAST-10
- If indicated by pre-screen: AUDIT
- NIH Toolbox Psychological Well Being (general life satisfaction and positive affect scales)
- PROMIS Physical Function
- PROMIS Social Participation
- PROMIS Anxiety
- NIH Cognition and Motor Batteries (in-person assessments - may be foregone for participants living in remote locations)
- ATHF
- CIRS-G
- Medical Conditions and Medication List (study doctor to review before randomization)
- MADRS

The key variables being used in the study include:

- Age
- Gender
- Race
- Ethnicity
- CIRS-G (cumulative medical comorbidity)
- MADRS
- PHQ-9
- ATHF

5.5.2. Descriptive summarization of baseline participant characteristics

We will list general patient characteristics in a baseline characteristics table. Data will be presented as mean with standard deviation (SD) when normally distributed or as median with interquartile range in case of skewed data. Dichotomous and categorical data will be presented in proportions.

6. Analysis

6.1. Outcome definitions

6.1.1. Primary and secondary outcomes

Primary: Psychological well-being is the patient-centered effectiveness outcome (H1a) and remission from depression is the clinician-focused effectiveness outcome (H1b). Safety will be monitored through serious adverse events (H2a) as well as falls and fall-related injuries (H2b).

Secondary: Secondary tests of effectiveness will examine changes in other aspects of quality of life: physical function, social participation, changes in depressive symptoms (i.e., MADRS scores), and antidepressant response (i.e., 50% or more reduction from baseline MADRS).

Continuation phase analyses (all are considered secondary):

In the continuation phase, we will explore the long-term effectiveness and safety of antidepressants:

- We will examine long-term changes/stability of psychological well-being and other quality of life variables, and depressive symptoms, by mixed-effects repeated measures models.
- We will also test for sustained remission (among those who achieved remission at the end of the acute step) by measuring relapse in each arm. Relapse will be defined as meeting DSM 5 criteria for a major depressive episode, as determined by an Independent Evaluator – who will identify all remitted participants who then have a criterion-based (ie PHQ9 score of 2-3 for items 1-8, 1-3 for item 9) and who will confirm the diagnosis by conferring with a study investigator. We will carry out Cox models examining time to relapse (primary: meeting PHQ-9 criteria for a current MDE; secondary: MADRS $\geqslant$15).and time to drop out, to examine for treatment inferiority (i.e., whether one antidepressant strategy produces poorer long-term outcomes).
- In the continuation phase, we will also explore the long-term safety of antidepressants. **Serious adverse events** are the safety outcome in both acute and continuation phases. Cox models examining time-to-event (with Anderson & Gill extensions for repeated events) will compare the treatment arms during continuation in terms of serious adverse events. Since falls and fall-related injuries information were not collected during the acute phase, the safety endpoint has changed. A time-to-event analysis will not be appropriated. Hence, a generalized linear model with logistic link function will be used.

We will also compute rates of side effects in each arm, as specifically recommended by patient stakeholders. These data are not hypothesis-testing, but they support the patient-centered goal of allowing clinicians and patients to know what to expect in terms of side effects from treatment.

As well, the study will test how aging influences the relative benefits and risks of antidepressants for TRD. Additionally, a qualitative study of patient and clinician partners will provide the lived experience of TRD and antidepressant strategies.

6.1.2. Outcome measurements

H1a outcome will be measured by NIH Toolbox (Psychological Well-being), and the secondary outcome will be measured by PROMIS (Physical function, Social participation).

H1b outcome will be assessed by independent (blind) raters via phone interview to assess the MADRS score.

H2a outcome will be measured when SAE occurs.

H2b outcome will be assessed by independent raters.

6.2. Analysis methods

To test hypothesis H1a (psychological well-being), a repeated measure of 2*2*3 ANOVA with age (60 – 70, > 70), time (baseline, end of Active phase of Step 1) by treatment group (3 levels) will be used. Specific time * treatment group contrasts will compare the changes across pairs of treatment groups. These tests will be conducted with a Hochberg Step-down procedure. If the comparison with the lowest p-value < 0.05/3 it is significant. If the second lowest p-value is also < 0.05/2 then it also will be significant and if the third p-value is < 0.05 then it also will be significant. We will include site as strata in all analyses.[5] For those who proceed to Step 2, analogous repeated measures 2*2*2 ANOVA will be used with a significance level of .05 for the time*treatment group comparison.

To test hypothesis H1b (remission), a corresponding 2*3 (age by treatment group) generalized linear models with a logit link function will be computed for the dichotomous outcome of remission. Pairwise treatment groups will be compared as in H1a.

To test hypothesis H2a (serious adverse events), Cox models will be used to examine time to event (with Anderson & Grill extensions for repeated events) will compare the treatment arms during both the acute phase and the continuation in terms of serious adverse events. Pairwise comparison between treatment groups will be tested.

To test hypothesis H2b (falls and fall-related injuries), a repeated measures generalized linear model with a cumulative logistic link function. The factors will be treatment group and time (week 2, 4, 6, 8 or 10). The age group (<70 vs.>70) and clinic stratification variables will be included in the model as well as the fall history obtained at baseline. Pairwise comparisons between the treatment group pairs will be formally tested. The same model with a simple logistic link function will be conducted during each 2-week period where the response is a fall-related injury or not.

In terms of the exploratory hypothesis 3, we will include the treatment by age group (<70 vs >70) by time interaction to the linear models described above to test for both the benefits and risks for the treatment alternatives. Additionally, we will fit a model with age as a continuous variable.

6.3. Missing data

Missing data can occur from patient dropout (unit non-response), failure to provide data (item non-response), or administrative issues. We will minimize missing data by closely monitoring accumulating data for missingness and implementing procedures to reduce it when identified (MD-1[6]). The distribution of patients with missing data for each variable will be thoroughly investigated (MD-1[5]). When a patient drops out of the study, we will document the reason for dropout and who made the dropout decision (MD-4[7]). We will compare patients with complete data to those with missing data to assess potential biases. Our approach will depend on both the pattern and the magnitude of the missing data, following strategies recommended by PCORI methodology standards and the National Research Council. For example, we

[5] Senn S. Testing for baseline balance in clinical trials. *Stat Med.* Sep 15 1994;13(17):1715-1726.
[6] Describe methods to prevent and monitor missing data.
[7] Describe how you will record and report all reasons for dropout and missing data, and account for all patients in reports.

will conduct sensitivity analyses (MD-5[8]) based on different assumptions (MD-2[9], MD-3[10]) including multiple imputations.

Our specific plan for handling missing data with the MADRS (at baseline or endpoint) is as follows: for those missing a centrally-administered MADRS, we will accept a locally administered one. For MADRS missing at baseline or at endpoint in completers, we will impute a MADRS score. For those who stop Step 1 (acute) or Step 2 (acute) early, prior to week 10, and do not have a MADRS score, we will treat them as non-remitters with respect to the primary effectiveness analysis.

6.5. Harms

Adverse events assessment is one of our research outcomes. During the study, regular assessments for side effects will be conducted. If present, they will be assessed by a physician. Similarly, we will make recommendations to clinicians that appropriate metabolic parameters (glucose, lipids), EKG, thyroid, renal, drug level, and other safety tests be monitored throughout the study. At any time during the study, if a participant is judged to be experiencing serious side effects, the research team will provide decision support so that clinicians can provide prompt management, including dose reduction, stopping treatment, or clinical referral.

6.6. Statistical software

Data cleaning and statistical analyses will be performed using SAS (SAS Institute, Cary NC) and R (The R Foundation for Statistical Computing; Vienna, Austria).

[8] Describe how you will examine sensitivity of inferences to missing data methods and assumptions, and incorporate into interpretation.
[9] Describe statistical methods to handle missing data.
[10] Describe plans to use validated methods to deal with missing data that properly account for statistical uncertainty due to missingness.