Pragmatic Obstructive Sleep Apnea Weight Loss Trial Assessing Effectiveness and Reach (POWER)

Data Analysis Plan v3.3

PI: Lucas Donovan, MD MS

Table of Contents

1. Abstract:

Proposed study design: POWER is a type 1 hybrid pragmatic RCT in which 696 Veterans with OSA and obesity are randomized to usual care or usual care enhanced with a remote and self-directed lifestyle intervention for 12 months. Our primary outcomes are change in sleep-related quality of life and weight at 12 months. We will assess secondary outcomes at 3-, 12- and 21-months.

Background: Prevalent obesity related conditions like obstructive sleep apnea (OSA) represent an important opportunity for the VA to improve population health. OSA markedly reduces quality of life and is associated with 3-fold greater risk for cardiovascular disease. Although obesity is the single greatest reversible risk factor for OSA, the 1 million Veterans with OSA and obesity rarely receive weight loss care to reverse OSA and other serious comorbidities. Efficacy trials reinforce that time and resource intensive lifestyle-based weight loss programs improve weight and physiologic measures of OSA severity (apnea hypopnea index, AHI). However, there are barriers to translating these findings into meaningful gains for population health. First, VA has limited capacity to counsel patients around lifestyle change. Less than one-third of Veterans with OSA and obesity are counseled about weight loss and even fewer are referred to weight loss services. Second, VA's current weight loss offerings are difficult to access. Only 12% of Veterans with OSA and obesity utilize MOVE! and those referred to MOVE! achieve minimal weight loss—1.2 kg at 1 year. Third, prior weight loss trials focused on intermediate measures (e.g., AHI), limiting understanding of effectiveness for meaningful outcomes. To meet these

challenges, we propose a pragmatic trial of proactively offering a remote video-based and self-directed weight loss intervention with telephone-based coaching to Veterans with newly diagnosed OSA. Our weight loss intervention (D-ELITE) is adapted from a program known to be effective in a non-VA population, with 44% of participants achieving ≥5% weight loss at 24 months. Further optimizing reach, our remote intervention includes low-technology options (e.g., videos on DVD and/or USB drives) to accommodate those with low technology literacy.

Significance: Our research tests a program of proactively providing Veterans with OSA the tools to manage weight loss in a way that is independent of local provider time and resources. Our research addresses a key gap in Veteran's health in a way that aligns with important VA priorities including population health, virtual care, access, and health care value. We anticipate our intervention can efficiently achieve improvements in quality of life while reducing the burden and risk of serious comorbidities.

Innovation and Impact: Our research directly challenges the traditional provider-driven model of healthcare delivery where providers direct care and provide necessary services aimed at managing a single disease. Instead, we propose to proactively deliver weight loss services to a high-risk group using a population health approach. In doing so, we will conduct the first trial of remote and self-directed weight loss care in OSA and will test whether weight loss care can improve meaningful outcomes such as quality of life and cardiovascular risk.

Specific Aims: Our primary aim is to test the effectiveness of a proactively delivered and pragmatic weight loss intervention to improve co-primary endpoints of sleep-related quality of life (Functional Outcomes of Sleep Questionnaire, FOSQ) and weight among Veterans with OSA and obesity. Secondarily, we will compare additional outcomes between groups: cardiovascular risk scores, sleep symptoms, and AHI. Finally, we will also conduct an implementation process evaluation informed by the RE-AIM framework to identify barriers and facilitators to widespread implementation.

Methodology: We plan a hybrid type 1 pragmatic randomized controlled trial. We will proactively identify Veterans with OSA and obesity nationwide using data from the CDW (n=696), randomizing 1:1 to usual care plus the D-ELITE weight loss intervention or usual care alone. We will collect primary outcomes at 12-months, but we will also collect outcomes at 3- and 24-months to assess trends over time. For our implementation process evaluation, we will use quantitative and qualitative methods to assess barriers to implementation, including a comprehensive budget impact analysis.

2. Sample Size Rationale:

We designed this pragmatic trial to achieve power of 0.9 for two primary outcomes at 12 months assuming 20% attrition using a two-sample t-test of independent means (STATA version 16.1). We target a two-sided type I alpha of 0.025 for each outcome to account for a Bonferroni correction. FOSQ: We plan to detect an established meaningful and important difference for FOSQ between groups, 0.75 points, and assume the known standard deviation for change in FOSQ of 2.5 points. Weight: Participants in a prior version of the D-ELITE self-directed intervention lost 4.5 kg. Assuming weight loss in our comparator group is similar to average Veterans with OSA seeking weight loss, 1.2 kg, we expect a 3.3 kg difference between groups. This difference approximates a meaningful difference of 3% weight loss (Hoerster et. al., JAMA, 2022. PMID: 36511927). Given known standard deviation in clinical weight change among Veterans' attempting weight loss of 6.9 kg/year, we will have 0.99 power to detect this 3.3 kg

difference in weight change with 20% attrition (Table 1). We will detect a 2.1 kg difference with power of 0.90. Across both co-primary outcomes, we retain power of 0.8 even with 39% or more attrition (Table 1).

3. Description of how data will be collected:

We will enroll 696 Veterans nationwide. In CDW, we will identify Veterans with an eligible outpatient BMI and sleep study recorded in the past 6 months with associated diagnosis of OSA who are free from exclusionary conditions searchable in CDW.

- **Research invitation**: We will invite potentially eligible patients by mail and email, and will send and invitation letter, information statement, and an opt-in/out postcard and Qualtrics link.
- **Telephone screen, consent, and baseline assessment**: Among those who do not opt-out, a research coordinator will review the patients' chart and will call to assess eligibility. Coordinators will describe the study in detail and obtain consent under a waiver of documentation of written informed consent. For those who consent, a research coordinator will administer baseline surveys over the telephone if possible.
- **Medical clearance:** After consent, but prior to randomization, we will send a secure email to each interested participant's primary care provider informing them of their patient's desire to participate. We will ask if their patient should not participate given the intervention content.

We will mail and email Veterans surveys at baseline and 3, 12, and 21-month follow-up. Research coordinators will administer surveys via telephone if they are not completed by mail or online through Qualtrics. We will assess baseline information using a health inventory checklist including socio-demographic characteristics, health behaviors, and non-VA weight loss programs. We will obtain additional information about comorbid conditions, medication use, healthcare utilization, and distance to clinics from the CDW.

Primary Outcomes**:** The co-primary endpoints for our pragmatic trial are changes in sleep-related quality of life and weight from baseline to 12 months. In secondary analyses, we will assess these outcomes at additional timepoints.

- **Sleep-Related Quality of Life:** FOSQ addresses domains that are important to patients: activity level, vigilance, intimacy/sexual relationships, productivity, and social outcome. FOSQ is sensitive to changes in OSA treatment and is widely used in trials of OSA management.
- **Weight Loss:** Consistent with the pragmatic nature of this trial, we will prioritize clinically measured weights using methods shown to be feasible and valid when compared to research weights. We will use participants' qualifying outpatient clinical weight prior to randomization as their baseline measure. To assess change, we will select the outpatient clinical weight that is closest to the 12-month post-randomization date within a 9–15-month post-randomization window, using a standard algorithm to remove implausible and erroneous values.

  Contingency for weight:
  We fully acknowledge that ongoing COVID-19 related changes in clinic utilization may alter the feasibility of the clinic-based approach. As a contingency, we also will request that participants weigh themselves with study-provided scales and report measured

weights shortly after randomization as well at 3-, 12-, and 21-month follow-up. Under a contingency approach that we will explore in secondary/sensitivity analyses, if a subject is missing a clinical weight, we will substitute self-collected weights in the analysis. We will provide written and verbal instructions to approximate clinic-based weights (e.g., clothed, but no shoes or jackets). Should we experience widespread clinic closures, limiting assessments of clinic weights, we may consider substituting clinic-based weights with self-weights in our primary analysis. We would make such a decision prior to analyzing primary outcomes.

Secondary outcomes: We will collect secondary outcomes including cardiovascular risk scores (non-laboratory Framingham), blood pressure, treatment adherence, and sleep symptoms (PROMIS measures) using a combination of surveys and CDW data. We will also mail home sleep apnea tests in a subset of patients at 12 month follow up only to understand impacts on physiologic severity of OSA.
All Veteran subjects will be remunerated $25 upon completion of the 12-month surveys, $25 for participation in the home sleep apnea test and $25 upon completion of any outcome surveys at the close of the study.

4. Method of Randomization:

The unit of randomization will be the patient, performed at a ratio of 1:1 within strata using the permuted-block method within block sizes of 4. Given three stratifying variables of BMI category (30-34.9 vs. 35-44.9 kg/m2), physiologic OSA severity (moderate to severe as defined by interpreting physician or apnea hypopnea index ≥15 events/hour vs. otherwise), and sleep related impairment (PROMIS Sleep Impairment T-score ≥ 60 vs. T-score <60) there will be 8 strata. Research assistants recruiting patients will remain blinded to randomization assignment sequences. Following baseline assessment, we will assign patients from each stratum to intervention or control using predetermined randomly generated sequences. Patients will be informed of their randomization assignment via mail. This will ensure a balanced number of patients in each stratum in the intervention and control groups. Stratification will help prevent covariate imbalance between groups.

5. Plans for any interim analysis of the data:

We do not plan to perform any interim analyses for superiority, futility, or sample size re-estimation. The D-ELITE weight loss intervention is consistent with existing VA guideline-based care for obesity and obstructive sleep apnea. Our trial is therefore meant to promote recommended medical care, and we do not believe an interim analysis of the data is warranted.

6. Missing and Misclassified Data:

For our primary outcomes, we assume 20% attrition to retain power of 0.9. However, we expect to retain sufficient power, 0.8, even if we experience dropout of over 39% (Table 1). Our pragmatic trial prioritizes outcome collection from EMR, including JLV and CAPRI, to reduce participant burden and enhance generalizability. While we attempt to minimize missingness and misclassification through efforts to optimize follow up survey completion and validated approaches to decrease measurement error, we anticipate missing data. We originally considered complete case analyses, but ultimately chose an approach based on linear mixed models and maximum likelihood that is valid under the missing at random assumption. Linear mixed models impute missing outcome data implicitly within a hierarchical model, using

individuals' baseline, adjustment, and stratification values. Missing values of covariates will be imputed using the chained equations approach (e.g. MICE), using single imputation if missing rates are less than 5% and multiple imputation otherwise. We will carry out the recommended sensitivity analyses to the MAR assumption using methods based on pattern mixture models and imputation, by assuming a range of perturbations of imputed outcome values and assessing differences in model conclusions.

Table 1. Change in Functional Outcomes of Sleep Questionnaire (FOSQ)					
Power	Alpha	SD (score)	Difference (score)	N	Attrition (%)
0.90	0.025	2.5	0.75	556	20
0.80	0.025	2.5	0.75	426	39
Change in Clinically Measured Weight					
Power	Alpha	SD (kg)	Difference (kg)	N	Attrition (%)
0.99	0.025	6.9	3.3	556	20
0.90	0.025	6.9	3.3	220	68
0.80	0.025	6.9	3.3	170	76
0.90	0.025	6.9	2.1	556	20

7. List and definitions of covariates to be included in models:

As outlined in Section 9, our primary model will include a number of covariates. First, we will use each of the stratification variables as defined in Section 4, including BMI category, physiologic OSA severity on diagnostic sleep study, and sleep related impairment. We will also include adjustment variables "A" collected at baseline including patient's age, self-identified gender, self-identified race, Charlson comorbidity index and rurality. We will define gender and race based on self-identified characteristics collected in baseline surveys. We will use age documented in the electronic medical record at baseline, calculate Charlson comorbidity index using ICD-10 codes documented in VA electronic medical record in the year prior to randomization, and we will define rural residence by Rural-Urban Commuting Area (RUCA) codes, utilized by the VA Office of Rural Health to define rurality.

8. Methods for dealing with data transformations:

We do not anticipate the need to transform data elements prior to analysis.

9. Definitions of the analytic cohorts:

Intent to Treat: Primary analyses of treatment effects will be intent-to-treat, with participants analyzed in the groups to which they were randomized.

Hypothesis Testing:
H1 (**Primary Hypothesis**): Lifestyle intervention will lead to greater weight loss and improvement in sleep-related quality of life at 12 months, relative to usual care.
H2 a (Secondary Hypothesis): Lifestyle intervention will lead to greater improvement in secondary outcomes, measured at baseline and follow-up relative to usual care
H2b (**Secondary Hypothesis**): Self-directed lifestyle intervention will lead to more favorable values for outcomes only collected at follow up (e.g., AHI, global ratings of change).

Using intention to treat, we will test primary (H1) hypothesis separately for each outcome (weight and FOSQ) using linear mixed effects model

$$Y_{ij} = \beta_0 + \beta_1 T_j + \beta_2 X_i T_j + \beta_3 Q_i + \beta_4 A_i + u_i + \epsilon_{ij}$$

which models for each patient i their measure for each outcome (FOSQ and weight separately), incorporating time of measurement T_j (0 or 12 months), treatment by time interaction $X_i T_j$, stratification variables Q_i, and adjustment variables A_i including age, gender, race, Charlson comorbidity index, and rurality. The model will include a random normal intercept u_i for each patient and a random normal error term ϵ_{ij}, all assumed independent. The variances of the ϵ_{ij} will be assumed different at baseline and 12 months. Here, β_0 is the average baseline FOSQ/weight among the control and treatment groups combined (assumed equal due to randomization), β_1 is the expected change in FOSQ/weight over time in the control group, and β_2, the quantity of interest in hypothesis H1, is the expected difference in the change in FOSQ/weight over time between the treatment and control groups, or equivalently the expected difference between groups at 12 months. In secondary analyses, we will incorporate primary and secondary outcomes from each time point (baseline, 3, 12, and 21 months). We will use the same model to test secondary outcomes captured only at follow up, omitting baseline values.

Additional Analytic Subgroups: In addition to analyses of the entire sample using intent to treat, we will analyze primary and secondary outcomes according to model 1 and model 2 in the following groups:

1) Subgroups defined by each of the three pre-specified strata (see section 4).
2) Subgroups defined by gender.
3) Patients who are prescribed continuous positive airway pressure and use <4 hours per night in the 90 days before the outcome period.
4) The subgroup of patients who are prescribed continuous positive airway pressure and use ≥4 hours per night in the 90 days before the outcome period.

10. Adverse and Serious Adverse Event Monitoring:

To ensure unbiased determination across arms at 12- and 21-month follow-up, we will ask participants to complete a survey about potential adverse events (AE), serious adverse events (SAE) and unanticipated problems (UP) employing a body system-based assessment. We define SAEs as those resulting in hospitalization or death.

Dr. Donovan or another licensed independent provider with experience reviewing AEs in clinical trials, Elizabeth Mattox ARNP, will assess (using participant self-report and chart review) each AE for duration (start and stop dates and times), expectedness in the study population, severity, outcome, treatment, and relation to study activity for each patient. We will also perform chart reviews at 12 and 21 months to look for discharge summaries within the VA.

The following are expected adverse events in the POWER population of participants who have OSA, a high BMI, and adopt healthy eating and physical activity program:

- Gastrointestinal symptoms related to change in diet
- Musculoskeletal symptoms or injury resulting from increased physical activity, including increasing symptoms such as chest discomfort, shortness of breath, and leg cramping

- Development of weight and OSA associated medical disorders including diabetes, hypertension, liver disease, cardiovascular disease, cerebrovascular disease, arrythmias, asthma, COPD, clotting problems, and other lung related conditions
- Development or exacerbation of mental health conditions, the worsening of which is known to be associated with poor sleep, including depression, anxiety, posttraumatic stress disorder, and bipolar disorder
- Development or exacerbation of eating disorders including binge eating disorder, anorexia nervosa, and bulimia nervosa.
- Development of other conditions associated with unhealthy health behaviors, such as from tobacco and alcohol disorders (e.g., cancer)
- Age related illnesses, such as pneumonia, urinary tract, and skin infections
- Motor vehicle collision or other accidents arising from excessive sleepiness
- Death

Summary of Edits:

Version 2.0, 11/23/2021

- Minor format changes and typo corrections
- **Section 3. Description of how data will be collected:**
    - Added use of email for recruitment and online questionnaire collection (Qualtrics)
- **Section 4. Method of Randomization:**
    - Removed rural vs. urban strata
    - Redefined physiologic OSA severity strata
    - Added sleep related impairment strata
- Section 10.
    - Added Elizabeth Mattox ARNP as additional reviewer of AEs

Version 3.0 2/16/2023

- Minor format changes and clarification that this is a "lifestyle intervention"

- Section 2. Sample Size Rationale
    - While our proposed sample size has not changed, we now present power analyses for between group differences for change in weight focused on a difference in the range of 3% weight loss, which is more widely accepted as clinically meaningful (Hoerster et. al., JAMA. 2023. PMID: 36511927)
    - The target sample size of 696 is still required to identify a clinically meaningful difference in our co-primary endpoint of functional outcomes of sleep questionnaire.

Section 3. Description of how data will be collected:

- Revised to state that the co-primary outcome of weight will only include clinically measured weights identified from CDW, as contingencies around self-weight are not currently needed given the current state of clinic operations.
- Also clarified that self-weights are not measured at baseline, but rather shortly following randomization, when intervention scales are sent.
- Updated description of home sleep apnea testing to confirm that these tests will only be collected at 12 months, not at baseline as we originally planned (see changes to secondary outcome hypothesis 2b to accommodate).
- Clarified timing of participant reimbursement.

- Section 4. Method of Randomization:

    - Removed gender as a stratification variable, although we will retain as a subgroup analysis (see edits to section 7 and 9 as well).
        - Prior to the start of randomization in April 2023, our biostatistician had concerns that a large number of strata could lead to imbalance in randomization. Our biostatistician was particularly concerned about gender as a stratification variable given the relatively small number of women Veterans. In order to avoid imbalance, our biostatistician recommended that we limit strata to the three variables for which we anticipated a

relatively equal distribution between categories. We implemented this change prior to the start of randomization.

Section 6. Missing Data

- Explicitly stated that we will test our hypotheses using a complete case analysis.

Section 9. Definitions of the analytic cohorts:

- Revised language around hypotheses to specifically denote primary and secondary hypotheses.
- Added formal hypothesis (H2b) for comparing secondary outcomes between groups at a single time point. This is the case for apnea hypopnea index as research HSATs will only be collected at 12 months.

Section 10. Adverse and Serious Adverse Event Monitoring

- Clarified ambiguous language around review of adverse events. The prior version could have been interpreted as indicating the PI and another clinician would *both* review each adverse event. The current version has been corrected to describe that PI *or* another clinician will review each adverse event.
- In consultation with co-investigators and our local IRB of record, we are adding the development or exacerbation of eating disorders as an expected event given the high prevalence of these disorders within the Veteran population (Mitchell et. al. Psychol Assess, 2021, PMID: 34292003).

Version 3.1 7/31/2023

- Updates throughout changing secondary outcomes collection period for 24-months to 21-months. This change is made to allow uniform capture of longer-term outcomes in the funded timeframe of the study

Section 3. Description of how data will be collected:

- Clarified our methods around collection of weight data
- Updated our approach to participant remuneration, including additional payments made after primary outcome collection at 12 months.

Section 6. Missing and Misclassified Data and Section 9. Definitions of the analytic cohorts:

In consultation with our biostatisticians, we reconsidered our prior approach of ANCOVA with "complete case analysis". Instead, we will pursue our primary and secondary hypotheses using linear mixed effects model because it is valid under the missing at random assumption. In addition, we will pursue multiple imputation using chained equations for missing covariate data if needed. (Note that outcomes have not yet been analyzed, and we are still in the process of collecting primary 12 month outcomes).

Version 3.2 4/15/2024

- Section 7. Updated to include definitions of the adjustment variables (age, gender, race, Charlson Score, rurality) which were previously included in Section 9 of our data analysis plan and published in our protocol paper (Donovan et. al. Contemp Clin Trials. 2023 Dec:135:107378. doi: 10.1016/j.cct.2023.107378).

Version 3.3 06/12/2024

- Section 7. Updated this section to clarify that we will use all stratification variables, already defined in section 4 and listed in our protocol paper, in our primary analytic model.
- Section 9. Corrected formatting/font of the analytic model.