Statistical Design and Power

Data analyses will proceed with the intent-to-treat sample. Initial analyses will compare groups on baseline and demographic variables using chi-square test (for categorical variables) and analysis of variance (for continuous variables). Variables that significantly differ between groups or are theoretically important, such as pain and other (non-opioid) substance use disorder will be used as covariates in multivariate analysis, when appropriate. For end-of-treatment outcomes, analysis will compare groups via chi-square, Kruskal-Wallis test, and analysis of variance (ANOVA) or analysis of covariance (ANCOVA) tests, as appropriate. Evaluation of changes in outcomes through the follow-up period will occur with hierarchical mixed models specified for dichotomous or continuous outcomes (PROC MIXED and PROC GLIMMIX SAS procedures). These analyses allow for intra-subject serial correlation, unequal variance and covariance structures, and unequally spaced assessment intervals. They also account for missing data by using all available data and maximum likelihood estimation of parameters. To use the mixed effect modeling analysis in assessing intervention effect, an interaction term (intervention group by time) will be included together with other covariates (e.g., age, ethnicity, chronic pain, anxiety, trauma history, utilization of psychiatric or medical services). A significant beta coefficient of the interaction term at $p < .05$ level will be used as evidence supporting the intervention effect. Analysis of change scores can analyze variables that cannot be normalized. For all analyses, effect sizes will be estimated, with parametric and non-parametric statistics as appropriate; effect sizes for effects by sex will be estimated. Statistical significance will be at $p < .05$ and analyses will occur via IBM® SPSS® Statistics V21 and SAS 9.4 statistical software package (SAS Institute Inc, Cary, NC).

Phase 1, Usability Testing Stage. Analyses will center around assessing usability, acceptability, and needs for program updates based on users' experiences. We will apply descriptive statistics to assess participants' comments and questions related to specific program/equipment features and global assessments. Usability outcomes will be: 1) ability to use the program and equipment without staff assistance after training (yes/no), 2) time (in minutes) to complete the CM process once trained, and 3) total number of deviations from the direct path of completing CM procedures after training. Acceptability outcome will be an average SUS score of $\geq$ 80.[132] Pearson product moment and Spearman's rank correlations will examine relations between baseline characteristics (e.g., reading level) and usability and acceptability outcomes. A MySQL database administered through phpMyAdmin will track and store usage statistics, including time on each page, time in app, frequency and timing of pulling instructions. Google Analytics will constitute a secondary source of statistics.

Phase 1, Proof-of-concept usability feld test Stage. Primary analyses will assess usability and acceptability, mirroring those above (Focus groups section). In addition, a MySQL database administered through phpMyAdmin will track and store usage statistics, including time on each page, time in app, frequency and timing of pulling instructions. Google Analytics will constitute a secondary source of statistics. SUS and program feature ratings will be examined with descriptive statistics.

Phase 2, RCT. will proceed with the intent-to-treat sample. Initial analyses will compare groups on baseline and demographic variables using chi-square test (for categorical variables) and analysis of variance (for continuous variables). Variables that significantly differ between groups or are theoretically important, such as pain and other (non-opioid) substance use disorder will be used as covariates in multivariate analysis, when appropriate. For end-of-treatment outcomes, analysis will compare groups via chi-square, Kruskal-Wallis test, and analysis of variance (ANOVA) or analysis of covariance (ANCOVA) tests, as appropriate. Evaluation of changes in outcomes through the follow-up period will occur with hierarchical mixed models specified for dichotomous or continuous outcomes. To use the mixed effect modeling analysis in assessing intervention effect, an interaction term (intervention group by time) will be included together with other covariates (e.g., age, ethnicity, chronic pain, anxiety, trauma history, utilization of psychiatric or medical services). A significant beta coefficient of the interaction term at $p < .05$ level will be used as evidence supporting the intervention effect. Analysis of change scores can analyze variables that cannot be normalized.

The Primary Aim 1 is to assess the efficacy of OARS+CM in promoting Suboxone treatment initiation. The primary outcomes will be the proportion of participants who schedule a Suboxone treatment intake visit, and the proportion of participants who complete the Suboxone intake visit to start outpatient treatment. These outcomes are usually non-normally distributed and will be analyzed accordingly. We will assess the effect of the intervention using generalized linear mixed model, controlling for the dependence of repeated measures and the clustering effects.

Sample size justification. We use the proportion of participants who complete the Suboxone intake to start treatment as the calculating indicator. Based on the findings from a large randomized clinical trial on MAT treatment for opioid dependence after an acute care visit,[158] we estimate the proportion of patients who will initiate MAT at baseline is about 35%. We hypothesize that OARS will increase the rate by 25% and OARS+CM will increase the rate by 50%. Using G-Power program (proportions procedure) with α=0.05, *N*=150 (50 each group), the calculated statistical power is >0.80 for the comparisons between TAU and OARS, TAU and OARS+CM, and OARS and OARS+CM.

Aim 2 is to assess the efficacy of OARS+CM for increasing Suboxone treatment adherence and abstinence outcomes. Main outcomes will be the percentage of opioid-free urine toxicology tests, and the percentage of group/individual therapy sessions attended. Secondary outcomes will be the longest duration of abstinence (LDA) outcome, measured at month 1, month 3 (end of study period) and month 6 (end of follow-up period), and sustained abstinence from Suboxone start through each follow-up. LDA and sustained abstinence will be as a composite of the greatest number of days with consecutive negative urine samples and self-reported no use; a composite of biochemical evidence and subjective report is needed because urine tests are too infrequent in the clinical setting to rigorously determine sustained abstinence based solely on biochemical evidence. Sustained abstinence and LDA are usually non-normally distributed and will be examined accordingly. We will assess the effect of the intervention using linear mixed model, controlling for the dependence of repeated measures and the clustering effects.

Aim 3 is to evaluate costs. Economic evaluation of the CM intervention is highly relevant to considerations of implementation and will be informed by expert guidelines.[159] The primary hypothesis is that the incremental cost-effectiveness (cost per participant who does not have repeat OUD-related emergency acute care, measured at the end of treatment and the last follow-up) will be acceptable and support sustainability. The study will also determine the incremental cost of the CM intervention per OUD patient treated, to identify the budget impact of adoption on a large-scale basis if the intervention is efficacious. For analyses, we will determine the Incremental Cost Effectiveness Ratio (ICER) in dollars per repeat acute care visit averted. This ratio is the difference in cost incurred by the OARS+CM group divided by the difference in the proportion of participants without repeat acute care services over the short (intervention period) and long term (last follow-up). The statistical uncertainty of the incremental cost-effectiveness will be estimated by bootstrap analysis of trial data, a method that preserves the covariance between costs and outcomes.[160] A cost-effectiveness acceptability curve will be plotted with the proportion of replicates that are cost-effective at different threshold values of willingness to pay per successful quit participant.[161] The *p*-value for the statistical test that the OARS+CM intervention is cost-effective will be found by determining the proportion of replicates that are above each threshold value. We will conduct a literature review of cost-effectiveness studies to identify likely threshold values of dollars per OUD person treated in general. Costs of activities that are solely for research purposes (weekly surveys, follow-ups) will be excluded. Also see the Commercialization Plan for our approach to cost-savings and cost-avoidance.