Mobile Enhanced Prevention Support for People Leaving Jail Statistical Analysis Plan

PI Nina Harawa

NCT 04036396

Mar 9, 2023

MEPS Outcomes and Statistical Data Analysis

Outcomes.

Outcomes are indicators of achieving various milestones. Specific Aim 1 outcomes are measures of PrEP uptake: (a) primary care provider who can prescribe PrEP, (b) screening/evaluation for PrEP, (c) begin PrEP, (d) adhere to PrEP, (e) remain on PrEP for at least 3 months. Specific Aim 2 outcomes are whether study participants have sufficient preventive screening, (a) HIV screening every 3 months, (b) STI screening every 6 months, and (c) at least one HCV test. Specific Aim 3's outcome is that each participant be enrolled in appropriate SUDs treatment: (a) received any SUD treatment post-release; (b) attended some SUD treatment services appropriate to ASAM level; (c) met at least 70% of recommended treatment activities/frequencies for ASAM level of care. Outcomes of the secondary aims are (i) obtaining appropriate follow-up care for those who test positive for HIV, STI and HCV; (ii) lack of recidivism; and (iii) service utilization patterns. For all outcomes except the last, our primary hypothesis is that the control group will do worse, and our secondary hypothesis is that App+Incentives+PM arm will be more effective than the App+Incentives.

Modeling Outcomes.

Each outcome will be analyzed in longitudinal intention to treat using analyses of all available data from Baseline and 3-, 6-, and 12-month follow-up interviews, mobile app, and abstracted record data on utilization, as appropriate. For PrEP cascade outcomes from aim 1, the model is a logistic generalized linear mixed model (GLMM) with a random intercept for participants and a binary 0-1 outcome with 1 indicating success (screening for PrEP say) and 0 indicating a lack of success (not screening for PrEP). We generally do not expect participants to have any PrEP-related positive outcomes at baseline, and we anticipate using only follow-up data at 3, 6 and 12 months in the analysis of specific Aim 1 outcomes. The fixed effects will be a mean for every date (3, 6, 12 months) for each intervention group, for a total of 9 parameters. The first hypothesis is a test for each outcome of differences at 6 months; secondary tests are for differences at 3 months and at 12 months. If the three groups are different at a given time point, we will do pairwise comparisons among the treatments, particularly between the two intervention arms with and without PMs. Specific Aim 3 is similar, though there may be some participants with appropriate SUDs treatment prior to baseline, thus we will include the baseline observation as part of the outcome. At baseline, groups have not been affected by interventions, thus all groups will be the same, so there is one additional parameter for the baseline mean. Otherwise, the analysis is similar to Aim 1 outcome analyses.

For specific Aim 2, the question in analyzing outcomes is whether the intensity (mean count) of health care visits per unit time is the same among the three groups. Thus, the analysis is a Poisson GLMM with log link and an offset equal to the log of the time covered by each observation, which models rates of visits per unit time. The advantage of this over binary logit GLMM modeling of 1=yes=had at least one visit in the past three months versus 0=no=no visits past three months, is that the binary model penalizes or rewards participants inappropriately for going a few days longer or shorter than 90 days in between visits. The parameter structure and statistical hypothesis testing is the same as for aim 1, though we estimate rate ratios instead of odds ratios. We will estimate the difference between the desired rate of visits (one per 3 months or per 6 months or per year) and the actual observed rate in each intervention arm.

Initial data analysis includes checking for out-of-range data and data cleaning. We will plot average outcomes and error bars as plus or minus two standard errors of the mean as a function of visit, broken out by intervention arm. These are called empirical summary plots.[1] Weiss (2005) describes our general approach to modeling longitudinal data as will be generated by this study. A GLMM random intercept model can be fit in the SAS software Proc Mixed (Cary, NC 2017); if there are problems fitting the data in Proc Mixed, we will move to Bayesian software such as JAGS[2] in R.[3]

Modeling Issues and Sensitivity Analyses.

Sensitivity analyses will include predictors that are predictive of missing visits or whose average levels differ across study arms. Predictors predictive of missingness are likely to occur; satisfactory randomization should prevent average predictor levels from differing across intervention groups. We will analyze missing visits as a binary outcome with logistic random intercept regression model as functions of key (baseline; time-fixed) predictors (for example: Baseline behavior (e.g. whether they had tested for HIV in the 3 months prior to baseline), age under 30 vs. 30+; type of SUD; ASAM Level I & II vs. III & IV; stimulant user vs. not; race/ethnicity; preferred substance, education level), prior to performing the final analysis. Variables that are significant predictors of missing visits will be included as predictors in sensitivity analyses of intervention effects. If substantial missingness occurs in key baseline predictors, we will use multiple imputation[4,5] to fill in the missing predictors. The missing visit analysis omits the baseline time point, as the baseline observation is required. To evaluate randomization, we analyze the key baseline predictors as outcomes in a one-way ANOVA. If some of the predictors differ by intervention arm, they will be included in sensitivity analyses.

Exploratory subgroup analyses.

Subgroup analyses will be conducted to understand how the intervention works. We will consider all of our predictors listed in the previous paragraph first as main effects and then as moderators to understand how predictors affect outcomes and how intervention effects possibly differ by groups defined by key predictors. For example, we will examine whether the intervention works well for meth users, but not for non-meth-users or works better for one race/ethnicity group over another. We will plot empirical summary plots (defined above) for all predictors for outcomes to understand predictor main effects on outcomes prior to subgroup analysis. Thus, for example, we will test for main effects for meth use and if present, we will attempt a multivariable model with meth-by-intervention group interactions.

Power Calculations.

For each of the 11 primary outcomes, we calculated power given the Principal Investigator's (PI) estimates of probabilities of success for the 6-month outcomes for each intervention arm. These estimates were based on the existing literature and on preliminary data from the Passport to Wellness study. The PI often specified ranges, in which case we calculated power for the low end and the high end of the ranges resulting in 16 power calculations. Power for the full sample of 399 for testing *H0: all arms have the same probability of success,* with a chi-square with 2 degrees of freedom (df) was almost always 1 to three decimal points for alpha=.05; the lowest power was .958 for probabilities of success for the three arms of (0.025, 0.075. 0.15). Some other examples of triples of probabilities we considered were (0.2, 0.6, 0.8), (0.2, 0.45, 0.55), (0.3, 0.5, 0.65). However, we expect a fair amount of loss to follow-up, so we also calculated power for 25% attrition, leaving 100 per group. Again, power is strong, but the lowest power for testing this H0 is now 86.

Second, we also considered power for comparing between the two intervention groups for testing *H0: both intervention arms have the same probability of success*: across a wide range of baseline probabilities, we have 0.8 power at alpha=.05 to detect a difference of 0.17 at n=133, and a difference of 0.20 at n=100 (accounting for loss to follow up) in each group. Because we will be fitting longitudinal models with more data than what goes into a chi-square test, we will have increased power over this basic chi-square test on 2 df.

Multiple Comparisons.

We have listed 11 primary outcomes at 6 months for the three specific aims. To adjust for multiple comparisons, we propose to use a new methodology proposed in Harwood, Weiss, and Comulada (2017) (HWC17).[6] For n=11 primary outcomes, we expect to reject on average 0.55 = 11*0.05 null hypotheses, even in the case of no difference between arms. The HWC17 methodology uses a correlated Bernoulli test of H0: p=0.05 for n=11 outcomes that rejects the null hypothesis of no effect when 3 or more of the 11 tests are significant at p=0.05. Rejecting this null hypothesis means that the interventions are not the same, gives a type I error of .05, and accommodates correlations across outcomes. We then declare individual outcomes that produced a p-value below 0.05 as significantly different across arms as well. We will separately test outcomes at 3 months, 6 months, and 12 months using this methodology.

1. Weiss RE. *Modeling Longitudinal Data*. 2005 edition ed. Springer; 2005:454.
2. Cunningham WE, Mosen DM, Hays RD, Andersen RM, Shapiro MF. Access to community-based services services and number of hospitalizations among patients with HIV disease: Are they related? *J Acq Immun Def Synd*. Dec 1 1996;13(4):327-335.
3. *A language and environment for statistical computing. R Foundation for Statistical Computing*. 2017:ISBN 3-900051-07-0. http://www.r-project.org/
4. Rubin DB. Inference and missing data. *Biometrika*. 1976:581-592.
5. Little RJA, Rubin DB, Little RJA, Rubin DB. Introduction. *Statistical Analysis with Missing Data*. John Wiley & Sons, Inc.; 2002:1-23.
6. Harwood JM, Weiss RE, Comulada WS. Beyond the Primary Endpoint Paradigm: A Test of Intervention Effect in HIV Behavioral Intervention Trials with Numerous Correlated Outcomes. *Prev Sci*. Apr 22 2017;doi:10.1007/s11121-017-0788-y