POSITIVE HEALTH CHECK EVALUATION TRIAL

Statistical Analysis Plan

NCT Number: NCT03292913

Date: 03/03/22

Data Analyses

Descriptive statistics—such as counts (percentages), means (standard deviations), and medians (interquartile ranges)—will be computed to summarize outcomes for participants in the intervention condition and the standard of care condition. For the primary outcome, the number and proportion of trial participants will be computed for the intervention condition and the standard of care condition who achieve VL suppression at the 12-month endpoint. Effect size measures, including Cohen's d for continuous outcomes and risk differences and relative risks for binary outcomes, with 95% confidence intervals, will be used to quantify the size of the difference in outcomes between the two conditions. Statistical significance of differences between binary outcomes for the two conditions will be tested by chi-square tests or by Fisher's exact tests, depending on sample size. Differences between conceptually continuous outcomes will be tested by *t*-tests (after an appropriate transformation when needed) or by nonparametric tests (e.g., median tests).

For each outcome, generalized linear mixed models (GLMMs) will be fit, using software such as PROC GLIMMIX in SAS software (version 9.4), to compare outcomes between the intervention condition and the standard of care condition, controlling for sociodemographic characteristics and trial site.

The appropriate form of the model will be selected based on the distributions of the variables, such as a logistic model for categorical variables (e.g., VL suppression) or a Poisson model for count variables (e.g., number of missed appointments). Similar models will be used to examine changes in VL values from baseline through follow-up while accounting for demographics and repeated measurements over time. A time by trial condition interaction will be tested within these models to determine if changes in VL values vary across the two trial conditions (i.e., Positive Health Check+ Standard of Care (PHC) vs. standard of care). In addition, time-to event methods, including Kaplan-Meier methods, Turnbull nonparametric interval censored estimation methods, and Cox proportional hazards regression models,

will be used to compare the time to first VL suppression between the two study conditions while controlling for demographic characteristics.

We will test for potential modifiers of the effect of the PHC intervention on the trial's primary outcome, namely VL suppression. Interaction terms will be tested in the regression models to assess differences in the effect of PHC by variables including age, sex, race, ethnicity, income, and clinics' standard of care (clinic level), followed by subgroup analyses to estimate intervention impact among those groups. In addition to the primary analyses that will compare outcomes among the PHC and standard of care groups, secondary analyses will be conducted to examine a dose response effect of the intervention by comparing outcomes among participants based on the number of times they used the PHC tool (0, 1, or 2+ times). Also, similar analyses will be conducted to quantify the impact of specific aspects of the intervention (e.g., particular PHC modules viewed).

Participants will be included in the analyses on an intent-to-treat basis. The extent of missing data will be examined at both the item-level and the person-level. If data appear to be missing at random, multiple imputation will be used for item-level missing data on explanatory variables, such as sociodemographic characteristics. In the case of person-level missing data (e.g., attrition), demographic characteristics of participants who have a final outcome measurement will be compared with participants who are missing data on the final outcome measurement. Characteristics that vary based on attrition will be included as control variables in the mixed effect regression models. Additionally, sensitivity analyses will be conducted to assess the impact of missing data imputation methods for the final outcome measurement, such as carrying the last measurement forward.

Sample Size Estimation

We calculated the target sample size by hypothesizing that 50% of the standard of care arm will have VL suppression at 12 months and that 62% of intervention arm participants will have VL suppression at 12 months, a difference of 12 percentage points. We planned to use Fisher's exact test at

the 5% significance level to detect a difference between proportions. With a linear adjustment for 25% attrition in each arm, we will need a sample size of 758 = 568/(1 – 0.25) to have 80% power to detect a difference between the standard of care and intervention arms.