ADOLESCENT MEDICINE TRIALS NETWORK FOR HIV/AIDS INTERVENTIONS

Statistical Analysis Plan

Adaptive Antiretroviral Therapy (ART) Adherence Interventions for Youth Living with Human Immunodeficiency Virus (HIV) through Text Messaging (SMS) and Cell Phone Support (CPS) Embedded within the Sequential Multiple Assignment Randomized Trial (SMART) Design

NCT03535337

Sponsor:	National Institute of Child Health and Human Development (NICHD) National Institute on Drug Abuse (NIDA) U19HD089875

Study Chair/Lead:	Marvin E. Belzer, M.D. Children's Hospital Los Angeles Los Angeles, CA, USA
Study Co-Chair/Co-Lead:	Karen MacDonell, Ph.D. Wayne State University Detroit, MI, USA
Recruitment & Enrollment Center:	Hunter College PRIDE Health Research Consortium
Analytic Core Analyst:	Samiran Ghosh, Ph.D. Wayne State University

Study Procedure Guide Version Date:	September 8, 2021

DATA MANAGEMENT AND ANALYSIS PLAN

1.1 Data Management and Data Quality

The first randomization involves allocation to CPS vs. SMS using a 1:1 allocation ratio. This step will use stratified randomization in permuted blocks to achieve balance in distribution of behaviorally and perinatally infected youth. The REC data team will program the randomizer in Qualtrics and will not participate either in the intervention delivery or in the adherence reinforcement steps. The protocol team will monitor race, ethnicity, and gender to ensure balanced distribution at the first stage. After 6 months, stratified randomization will occur with permuted blocks within each trajectory. Since second stage randomization is dependent on first stage response status, we will only maintain the balance within each first stage treatment outcome when randomizing at the second stage.

Following the Analytic Core's (AC) standard quality assurance, including checking for outliers and abnormal values using graphical methods, we will verify that the distributions of measures meet the assumptions of the statistical tests to be used, applying a formal test such as the Shapiro-Wilk's test [60]. Transformations will be used when distributional assumptions are not fulfilled. Tests will be conducted to identify potential relationships among baseline demographic and clinical variables and our dependent variables and to see whether they are balanced between groups. If a baseline variable is not balanced between groups and is correlated with the dependent variable ($r>.30$), we will include this variable as a covariate in subsequent analyses. Since we will test more than one primary hypothesis with one primary outcome variable (VL suppression), the Hochberg step-up multiplicity adjustment will be used with a two-tailed family wise alpha-level of .05 [61]. All other tests described below will each have a two-tailed alpha-level of 0.05. Outcome analyses will be based on the principle of intention-to-treat [62].

1.2 Quantitative Analysis Plan

Original Primary Hypotheses: 1. Youth randomized to CPS will have significantly greater VL suppression (primary outcome) and self-reported medication adherence (secondary outcome) at 24 weeks than those in SMS group. 2. Non-responders randomized to CPS with incentives will have significantly greater VL suppression (primary outcome) and self-reported adherence (secondary outcome) than those randomized to SMS with incentives.

Original Secondary Aims: a) Conduct an exploratory evaluation of tapering the intensity of CPS and SMS (among responders) vs termination after completion of the 24 week intervention to determine if this improves the durability of intervention effects; b) Explore the relative efficacy of the individual intervention sequences embedded within the SMART design for responders and non-responders; and c) Describe moderators of the treatment effect, including demographics, mode of infection, substance use, and mental health. [NOT COMPLETED DUE LOW RENTENTION]

Updated Primary Outcome: Originally, we plan to treat VL suppression as our primary outcome for both co-primary hypothesis. Primary end-point for H1 was at 24 weeks when the proximal outcome was used to determine responder status, while for H2 the Distal outcome was at end of 12 months (or 72 weeks). Due to the difficulty in collecting viral load we proposed to change the primary outcome to a combination of VL supression and self-reported adherence (SRA). Primary end-points for both co-primary hypothesis is now at 48 week.

Rationale of new definition of Primary Outcome: Since the VLS data may not be available for a significant number of subjects, we plan to use SRA as an additional primary outcome. The usefulness and appropriateness of the usage of these two depends upon the concordance of two measures. To determine that for those subjects we have both VLS and SRA data available, we will first construct a 2×2 contingency table and conduct a chi-square test of association. If they agree on 80% or more cases we will proceed to define (and analyze) the primary outcome as following:

1) For any data collection point and randomization event, if VLS data is available it will be used as primary outcome. When VLS data is not available, SRA will be used as primary outcome. This composite definition will be used as primary outcome and will have sufficient power based on hypothesized effect size.

2) In case it fails to agree at least 80% cases, we will analyze VLS and SRA as two separate primary outcomes. Two measures cannot be combined, as they do not match sufficiently. In this situation, it is likely that VLS as primary outcome will not have sufficient sample size to produce an appropriately powered analysis, albeit, SRA may still produce a powered analysis provided SRA is available for all subjects. Note though no conclusion about VLS can be drawn from the outcome of SRA analysis.

1.2.1 Primary Analysis Plan

The primary analysis will be a comparison of SRA (primary outcome) between the CPS and the SMS group at the end of first stage of randomization. This will be performed using a χ^2 test. We will also compare the VL supression between the two groups for those avaialble subjects, albeit this is no longer treated as primary outcome. Since SRA (and VL) is a continuous measure, we will use two-sample t-test to conduct this analysis. We will also compare the self-reported medication adherence rate (secondary outcome) between the two groups using a χ^2 test. All the primary analyses will be based on initial assignment to groups, using the ITT principle. Each of the primary hypotheses will be tested using linear mixed-effects (LME) regression analyses.[104] For testing primary hypothesis 1 (PH1), the model will include up to 3 repeated assessments of VL suppression (weeks 0/baseline, 12, 24, 36, 48). For primary hypothesis 2 (PH2), we will only focus on non-responders (for both CPS and SMS at stage 1) and compare the SRA as dependent variable using a repeated measure LME (weeks 0/baseline, 12, 24, 36, 48). In an effort to get the outcome for PH2 some subject may have final outcome at latter date than 48 weeks. Each LME model will include a random intercept and slope and fixed effects for adherence intervention group, and time, as well as the stratification variables: clinical site, age, gender. A likelihood ratio test will examine the incremental contribution of the group by time interaction, which represents the interaction of interest for PH1 and PH2, testing for a differential adherence intervention effect over time. The decision rule for each primary hypothesis calls for rejection of H_0 if this interaction is statistically significant using the Hochberg adjustment. A site by group interaction will be also examined and included in each model (above) if significant at the 0.05 level. The purpose of the LME based analysis for the primary aim is to determine which of the first-stage intervention, CPS or SMS, is associated with the most improvement in VL and adherence, regardless of which second-stage treatment participants received. Due to contraction of sample size compared to our original plan, we do not plan to adjust α for multiple of testing. In addition to PH1, for those subjects we have VL suppression data available we would perform similar LME to test for its significance. We also expect some subjects with complete VL data and many with partial VL data for

PH2. An exploratory analysis is also planned where we define a composite binary outcome as 1 if either SRA or VL suppression (VL < 200 copies/ML) is achieved and 0 otherwise. This binary composite outcome variable will be analyzed using mixed effect GLM. Since our modified design guarantees availability of at least one of the two defining quantities this exploratory analysis ensures usage of full data.

1.2.2 SMART Secondary Aims

There are 3 secondary aims. The first secondary aim is to compare the effect of tapering with termination at the 2nd randomization among those who received CPS as well as those who received SMS and achieved early response (SRA originally was VL <200). This will be done again first by a comparison of two SRA rates among the two groups (tapering vs. termination), followed by more refined analysis using LME modeling. The second secondary aim is to compare the effect of CPS and SMS tapering vs. termination among the responders. We will first perform a χ^2 test between the two viral suppression rates, followed by mixed-effects modeling. The purpose of the second, secondary aim is to determine which of the eight adaptive interventions embedded within the SMART (listed in **Section A.5** in original proposal) leads to the greatest improvement in VL and SRA over the entire study period (48 weeks). To perform this, we will estimate the viral suppression rates and adherence rates among the subjects following each of the eight embedded adaptive interventions, and conduct a χ^2 test. Since in the current study, both responders and non-responders are re-randomized, there is no need to use inverse-probability weighting[106] of the study sample. However, in order to account for the correlation induced by subjects shared between any two embedded adaptive interventions, we will use *robust (sandwich) standard errors* as in the *generalized estimating equations* approach.[107] The final secondary aim is to study the moderators of treatment effect. This is a very interesting goal, given the gradual but assured paradigm shift in behavioral interventions from "one-size-fits-all" approach to the modern personalized medicine. Potential moderators in the current context are subject's demographics and the level of motivation – these can be incorporated in the analysis of the SMART data to deeply personalize the adaptive intervention for future patients. Because of the two-stage nature of the adaptive interventions, unfortunately a straightforward regression analysis including potential moderators in the model as interaction terms is not suitable due to the possibility of 'collider-stratification bias'–a type of selection bias that can be present in time-varying settings, even in presence of randomization, e.g. in a SMART.[108] To avoid this kind of bias, one needs to employ 2 separate regressions corresponding to the 2 stages of SMART, and carefully move backward through the stages; such a state-of-the-art approach is known as Q-learning.[109] Each of these regressions will contain interaction terms between the stage-specific treatments and the appropriate stage-specific moderators. If any of these interactions come out significant, then the corresponding patient characteristics can be used to deeply tailor the interventions for future patients. This analysis will be performed using the R software package *qLearn* (http://cran.rproject.org/web/packages/qLearn/index.html). Secondary aims are not powered.

1.2.3 Site Clustering Effects

Although some of the participants will be recruited via a site referral, we do not anticipate substantial site effects. A group identifier for each participant will be included in the merged analytic dataset, and the intra-class correlation (ICC) for each outcome within sites will be calculated prior to conducting multivariate analysis. If no significant variance (<0.05) is carried at the group level, we will reduce the

model to a traditional two-level model (only clustering due to repeated measure). If significant group-level variance does emerge, dummy codes to control for site-specific variance will be used to enhance statistical power using a 3-level LME model. In other words, ICC will be included in the model at three-level to control for the clustering effect by design, provided third level is significant. The significance test is always at 0.05. We did not recommend a three-level model all the time as power analysis is conducted considering two-level models only i.e. ignoring site effect.

1.2.4 Sample Size and Power

Sample Size and Power. Statistical power analyses examined the sample size requirements to detect greater than chance group differences on primary outcome (PH1 and Ph2). We have used Susan's formula with a total sample size of 120.

http://people.seas.harvard.edu/~samurphy/papers/APPAproof.pdf

Power Table

ES	p	N	N(10% attrition adjustment)
0.55	0.4	86	96
	0.5	**104**	**116**
	0.6	130	144
	0.7	174	192
0.6	0.4	74	80
	0.5	88	98
	0.6	110	120
	0.7	146	160

Alternatively, for the two groups comparison this corresponds to n=60/group. This sample size is sufficient to provide 80% power (or more) for a standardized Effect Size (ES) of 0.55 or more, using a t-distribution for a two tailed test with α=0.05, with 10% attrition assuming equal responder rate (p) in both arms.

1.2.5 Strategies for Attrition

Attrition introduces bias and reduces power, precision and generalizability [71]. In keeping with the Intent-to-Treat principle we distinguish between intervention and study termination (complete withdrawal). Accordingly, we will make every effort to continue assessments for the entire course of randomized intervention, even among those who do not continue with randomized group [72]. The mixed-effects models will incorporate all available data, even from subjects who do not complete the study. Mixed-effects models yield valid inferences assuming ignorable attrition [73]. For conducting descriptive and two-group comparison Multiple Imputation will be used to replace missing values in the outcomes and other measures. Point estimates, standard errors, and all tests will be calculated using Rubin's rules for combining the results of identical analyses performed on each of the imputed data sets [74].

1.2.6 Additional Secondary Aims

Self-management model analysis will be conducted based on the exploration, preparation, implementation, sustainment (EPIS) model. For the implementation aim, we will explore provider perceptions of the centralized implementation intervention and of the intervention as an evidence-based

practice, through qualitative interviews and quantitative assessments based on the EPIS model. Data coding and analysis will proceed in a three-phase process with the EPIS Team. First, consistent with Morgan's [75] recommendations for qualitative content analyses and Hsieh and Shannon's [76] directed qualitative content analytic approach, standard definitions of the concepts to be coded in the text will initially be developed by the ISC based on the EPIS model. We will first systematically review each interview at each time point for all thematic mentions of EPIS model constructs, initially using existing theory to guide categorization, but also allowing themes to emerge from the data through open coding procedures [77, 78]. This combined inductive and deductive coding approach will allow us to both validate and extend the EPIS framework through our analysis. Revision of our initial coding categories will occur iteratively until we reach saturation in the identification of new codes. During this iterative process, categories and their definitions will be refined and sub-categories of codes consolidated, consistent with an axial-coding process. At this point, we will return to each interview and systematically apply the final, revised set of codes. In addition, case codes will be applied to each interview to reflect clinic role, site, cluster, and relevant demographic characteristics of the respondent. All coding will be conducted using NVIVO Version 10. For reliability, a random selection of 30% of the interviews will be independently coded. Coding will be monitored to maintain a kappa coefficient of .90 or higher [79, 80]. Any discrepancies will be discussed and resolved. We will work closely with the EPIS team and then ISC to develop an Intervention Profile and Implementation Resources for replication and sustainment of the intervention. The Profile will synthesize intervention components and implementation analyses into intervention-specific practical guidance for further scale up. Additional resources to facilitate uptake, quality implementation, and sustainment of effective interventions are described in the ISC.