Statistical Analysis Plan

Reducing HIV Vulnerability Through a Multilevel Life Skills Intervention for Adolescent Men: The iREACH Project

Principal Investigators: Rob Stephenson, PhD, University of Michigan
Jose Bauermeister, PhD, University of Pennsylvania
Patrick Sullivan PhD, Emory University

Protocol identification number: 1U01MD011274-01

ClinicalTrials.gov identifier: NCT03155841

Author: Rob Stephenson, PhD

Version: Draft 1.0

SAP Signatures

I give my approval for the attached SAP entitled Reducing HIV Vulnerability Through a Multilevel Life Skills Intervention for Adolescent Men: The iREACH Project, February 22, 2023, version 1.0

Statistician (Author)
Name: Rob Stephenson

Signature: Robert Stephenson

Date: 2/23/23

Statistician Reviewer (As applicable)
Name: NA

Signature: _____

Date: _____

Principal Investigator
Name: Rob Stephenson, PhD

Signature: Robert Stephenson

Date: 2/23/23

Principal Investigator
Name: Jose Bauermeister, PhD

Signature: _____

Date: 2/23/23

Principal Investigator
Name: Patrick Sullivan, PhD

Signature: _____

Date: 02/23/2023

Abbreviations and Definitions
AMSM: adolescent men who have sex with men
LGBTQ: lesbian, gay, bisexual, transgender, and queer
MSM: men who have sex with men
RCT: randomized controlled trial
SMART: Study Management and Retention Toolkit
SMS: short message service
STI: sexually transmitted infections

Introduction

Preface
Between 2000 and 2010, the annual number of new HIV diagnoses among young men who have sex with men (YMSM; age 13-24 years) in the United States more than doubled. In 2016, youth aged 13 to 24 years made up 21% of all new HIV diagnoses in the United States. Most (81%) of those new diagnoses occurred among young gay and bisexual men. Young black/African American and Hispanic/Latino gay and bisexual men were especially affected. The HIV epidemic among YMSM is characterized by strong racial and ethnic disparities in HIV incidence. Data from the National HIV Behavioral Survey showed 26% of African American youth surveyed (aged 18-24 years) tested HIV positive compared to only 3% of white youth aged 18 to 24 years. Although initially attributed to greater engagement in risk behaviors among racial/ethnic minorities, recent analyses illustrate that racial/ethnic disparities in HIV are likely driven by social determinants of health, including access to health insurance and social network properties, and by more limited coverage of effective HIV treatment for those living with HIV. These findings underscore the importance of understanding and addressing the structural factors driving the HIV disparities faced by marginalized YMSM in the United States. Few HIV interventions have demonstrated efficacy in reducing HIV risk among adolescent men who have sex with men (AMSM), and fewer still have recognized the unique needs of AMSM based on race/ethnicity or geographical setting. Recognizing that youths' HIV vulnerability is intricately tied to their development and social context, delivering life skills training during adolescence might delay the onset or reduce the consequences of risk factors for HIV acquisition and equip AMSM with the skills to navigate HIV prevention. This randomized controlled trial (RCT) aims to test the efficacy of an online-delivered life skills intervention, iREACH, on cognitive and behavioral HIV-related outcomes for AMSM.

Scope of the analyses
The primary objective of this randomized controlled trial (RCT) is to test the efficacy of an e-delivered life skills intervention, iREACH, on cognitive and behavioral HIV-related outcomes for AMSM. We will recruit a large and diverse sample of AMSM (N=600; ≤50% non-Hispanic white) living in four regions disproportionately burdened by HIV prevalence across the United States.

Study Objectives and Endpoints

Study Objectives

Aim 1: Adapt a multilevel, online life skills intervention to address HIV vulnerability among AMSM living in four heavily impacted regions constituting diverse racial/ethnic populations and US geographic areas (Chicago-Detroit; Atlanta-Washington, DC; Memphis-New Orleans; San Francisco-San Diego).

Aim 2: Test the efficacy of iREACH, as compared to a delayed intervention condition, to improve cognitive (e.g., comfort discussing sexuality; HIV prevention attitudes, norms, self-efficacy, behavioral intentions) and behavioral (e.g., condom use, HIV testing, PrEP use) factors using a prospective RCT design.

Aim 3: Examine the differential efficacy of iREACH in shaping the psychosocial mediators (e.g., personal competency) associated with our outcomes.

Aim 4: Examine how socioecological determinants at the individual (e.g., race/ethnicity, urbanity) and regional (e.g., socioeconomic disadvantage, HIV prevalence) level are associated with iREACH's efficacy.

Endpoints

Primary outcomes: changes in confidence to engage in HIV prevention behaviors and changes in HIV knowledge. **HIV knowledge is measured using a recently updated measure of the HIV-KQ for AMSM**, an internally consistent and stable HIV knowledge scale shown to be appropriate for low-literacy populations. **Motivation to engage in condom use behaviors** will be assessed using 3 scales from Fisher et al.'s Teen Health Survey. We will measure motivation using 3 scales: attitudes (3 items; α>.60), social norms regarding peers and partners (6 items; α>.80), and behavioral intentions (3 items; α>.80). **Motivation to adopt routine HIV testing** (e.g., making a calendar for HIV testing or talking to partners about HIV testing) will be assessed with 3 subscales assessing AMSM's attitudes, social norms, and behavioral intention that we have used in the past with this population. Each attitude item is measured with three five-point semantic differential scales (*good-bad*; *worthless-valuable*; *pleasant-unpleasant*). Social norms assess the extent to which participants felt that friends and family believed he or she should test for HIV on a 1 to 5 scale. Behavioral intentions items assess participants' intention to adopt HIV testing, on a 1 (*very unlikely*) to 5 (*very likely*) scale. We will measure **self-efficacy** across three behaviors: condom-use self-efficacy (3 items; α=.72), safer-sex negotiation self-efficacy with casual partners (5 items; α=.94) and regular partners (5 items; α=.91), and discussing HIV testing with partners[110] (3 items; α=.90),

Secondary outcomes: changes in sexual behavior and change in HIV testing behavior over the 12-month follow-up period. **Sexual risk behavior** will be assessed using the Sexual Practices Assessment Schedule for Youths. We have adapted this assessment for use in previous online studies with YMSM. The assessments ask lifetime of sexual partners and types of sexual acts (e.g., kissing, touching, oral sex, penetrative sex) before proceeding any further. For behaviors reported, the assessment then explores the number of occasions of different sexual acts (oral, anal; receptive, insertive) with three different types of partners (romantic interest, casual partner ["hookup"] or friend with benefits), use of condoms during the past 3 months, and knowledge about partners' HIV status. The assessment also differentiates between male and female partners. We will adapt the assessment to ascertain whether, to their knowledge, their sexual partners are on PrEP. **HIV/STI testing and care:** The baseline survey will include questions on lifetime HIV/STI testing history[271]. Follow-up surveys will repeat the questions from the baseline and will also include questions on HIV/STI testing in each 3-month period including test results. Among newly diagnosed HIV+ cases, we will assess their linkage and retention in care.

Study Methods

General Study Design and Plan

We will conduct a prospective RCT of 600 online-recruited cis-gender AMSM (age 13-18 years) followed for 12 months with study assessments at each 3-month interval. A racially/ethnically diverse sample (at least 50% racial/ethnic minority) of AMSM living in four regions in the United States: (1) Chicago, IL to Detroit, MI; (2) Washington, DC to Atlanta, GA; (3) San Francisco, CA to San Diego, CA; and (4) Memphis, TN to New Orleans, LA. Regions were identified by inspection of HIV prevalence rate maps on AIDSVu.org. Eligible counties are those that include the major interstate highway that connects the two anchor cities (ie, I-94 for Chicago to Detroit; I-95 for Washington, DC to Atlanta; I-5 for San Francisco to San Diego; and I-55 from Memphis to New Orleans). Each region includes urban, suburban, and rural counties, as classified by the 2006 National Center for Health Statistics urban-rural classification scheme for counties [15].

Inclusion-Exclusion Criteria and General Study Population

Eligible individuals must (1) have been assigned a male sex at birth and identify as male at the time of enrollment into the study (cis-gender male), (2) be between the ages of 13 and 18 years (inclusive), (3) speak and read English, (4) report same-sex attractions and/or behaviors, (5) have access to the Internet, (6) live in one of the zip codes at least partially contained in the 109 counties included in the four regions selected for this trial, and (7) self-report as HIV-negative at time of enrollment.

Randomization and Blinding

Participants are randomized on a 1:1 basis using a stratified randomization by race and region. AMSM will be randomized to one of two arms, with 300 randomized to each arm (complete data on 250 AMSM per arm, allowing for attrition of 50 AMSM per arm).

Study Assessments

The proposed domains mirror the theoretical framework informing our intervention. We include a summary of the primary and secondary outcomes, key intervention mechanisms of change, and potential covariates. Assessments will not take more than 25 minutes to complete. We have taken several steps to ensure that the assessments are brief and developmentally appropriate. First, AMSM will not answer all listed constructs at every assessment. Identity-related constructs, for example, will be measured at 6-month intervals, whereas other constructs (e.g., comfort discussing sexuality) will be assessed at each follow-up. Second, consistent with best practices that we have used in prior HIV studies with pediatric and adolescent populations[142,256], we will use gate-keeping questions to decide whether participants should receive other probes within a domain. For example, a participant who has yet to become sexually active would receive cognitive assessments focused on HIV risk and awareness (e.g., attitudes, norms), but would not receive questions regarding behavioral intentions to engage in HIV prevention in next 3 months (e.g., HIV testing, PrEP) or questions regarding sexual risk behaviors (e.g., sexual behavior, HIV testing, PrEP use). Similarly, at each assessment, AMSM with no lifetime substance use will not be asked about additional substance use questions. Among sexually active participants, gatekeeping also provides opportunities to regulate how detailed the assessments will be.

Domain	Assessment time					
	Baseline	Month 3	Month 6	Month 9	Month 12	Month 15
Primary outcomes						
HIV knowledge	X		X		X	X
Condom use/communication efficacy	X		X		X	X
HIV/STI testing	X		X		X	X
Outcomes required for analysis						
Demographics	X		X		X	X
Patient provider communication around sexual orientation		X		X		
Stigma		X		X		
Resilience		X		X		
Psychological needs	X		X		X	X
Future life goals		X		X		
Secondary outcomes						
Internalized homonegativity	X		X		X	X
PrEP use and willingness [	X		X		X	X
Sex behaviors	X		X		X	X
Substance abuse	X		X		X	X
Depression	X		X		X	X
Anxiety	X		X		X	X
Self-Esteem	X		X		X	X
Covariates						
Peer influence		X		X		
Family support		X		X		
Discrimination		X		X		
Online behaviors		X		X		
Societal reaction to sexual orientation		X		X		
Ethnicity beliefs		X		X		
Relationship history	X		X		X	X
Intervention acceptability		X				X

Sample Size

Our expected sample size for analyses across both conditions is N=600 (Intervention, n=300; Control, n=300), assuming a 15-20% loss to follow-up. We estimated the minimum detectable effect sizes at 80% power, for comparisons of the two groups for the continuous primary cognitive (e.g., behavioral intentions, attitudes, norms, self-efficacy) outcomes, as well as the difference in the proportion of AMSM who report increased behavioral in the outcomes (e.g., condom use, HIV testing) to engage in HIV prevention services and behaviors in our treatment arm vs. our attention-control arm. For mean differences, our sample size calculations are based on a two-sample t-test assuming equal variance using a two-sided significance of 0.05. At 80% power, we can detect a between-arm difference of d=0.22 at the final follow-up. For repeated measure analyses, assuming a within-person correlation of 0.25, we would be able to detect a difference of 0.08. A less favorable within-person correlation of 0.75 allows us to detect an effect size of 0.11. For proportions, our sample size calculations are based on a two-sample test of proportions using a two-sided significance of 0.05. To have 80% power to compare active treatment to the control group, we require at least 500 participants to find a 12.5% difference between treatment and control in cross-sectional analyses. Assuming within-person correlation of 0.25, we can detect an 8.8% difference. A less favorable within-person correlation of 0.75 allows us to detect an 11.3% difference.

Timing of Analyses
Final analysis will be conducted once the desired sample size is recruited has reached and completed the last study assessment.

Analysis Populations
Intention-to-treat (ITT) All randomized study subjects. This will be seen as the primary population for the analysis. Per Protocol (PP) All randomized study subjects completing the whole study period (complete cases). For a specific analysis, study subjects with missing data on any of the variables in the model will be excluded from the analysis. Analyses of this population is seen as a sensitivity analysis to investigate whether conclusions are sensitive to assumptions regarding the pattern of missing data.

Covariates and Subgroups
We will also collect information of the following mediators and moderators for our analyses: Socio-demographic information will include questions on race/ethnicity, educational attainment, employment status, place of birth, year of immigration (when appropriate), housing status, history of incarceration, and engagement in transactional sex. Psychological distress will be measured using existing, well-validated scales. We will use subscales from the Brief Symptom Inventory to measure depressive symptoms and anxiety symptoms. Depressive symptoms include 6 items (α>.80) rated on a 5-point scale (1=never, 5 = very often). Items include feelings of loneliness, blue or sad, and having thoughts of ending one's life. Anxiety also includes 6 items (α>.90) measured on the same 5-point scale and includes reporting nervousness or shakiness, feeling fearful, or suddenly scared for no reason. We will also measure AMSM's Perceived Loneliness using the UCLA Loneliness Scale using a 4-point scale (1=Never;4=Often). In our prior work, reliability for this scale with YMSM was high (α>.90). To measure Stress, we will assess perceived stress through daily hassles. AMSM will rate their experiences of stress over the past month (e.g., how often have you found that you could not deal with all the things that you had to do), and rate the frequency with which they felt these ways on a scale from 1 (Never) to 5 (Very often). We have used these measures in prior studies and have shown excellent psychometric properties consistently over time. Social Media Use: Given potential disparities in access to technology, we will include the PEW Internet Survey questions regarding use of different devices, the number of hours spent online through each device, the reasons for social media use, sites commonly frequented, and extent to which the Internet supplements face-to-face interactions. We will also measure AMSM's frequency of use of social media to look for HIV or sexual health-related information, and their online partner-seeking behaviors. Online health seeking competency will be measured using an adapted, 6-item version of the eHealth Literacy Scale (eHEALS) that asks respondents on a 5-point Likert scale about their knowledge of online resources including where to locate them, how to find them, how to use information, and comfort with assessing the quality of online health information.

Missing Data
To allow for 15-20% loss to follow-up (our previous trials have each achieved retention rates of >90%), we estimate a sample of 600 AMSM to be enrolled across the four regions. Participants may continue the study even if they miss surveys intermittently over the data collection period. We will compare those who completed different follow-up interviews with those who did not on key predictors from the baseline assessment to check for possible sampling bias due to missing data. Missing data will be minimized by the web-based entry for all measures. We will handle missing data by applying likelihood-based methods that include partial information in cases that drop out prematurely and develop nonresponse adjustment and post-stratification weights to account for survey non-participation. The use of Expectation-Maximization (EM) algorithm and multiple imputation approach

in longitudinal analyses will help overcome missing data concerns when appropriate. When the missing at random assumption appears suspect, we will not replace missing data to avoid bias.

Summary of Study Data

All continuous variables will be summarized using the following descriptive statistics: n (non-missing sample size), mean, standard deviation, median, maximum, and minimum. The frequency and percentages (based on the non-missing sample size) of observed levels will be reported for all categorical measures. In general, all data will be listed, sorted by site, treatment, and subject, and when appropriate by visit number within subject. All summary tables will be structured with a column for each treatment in the order (Control, Experimental) and will be annotated with the total population size relevant to that table/treatment, including any missing observations.

Efficacy Analyses

Analysis tests the efficacy of the iREACH intervention as compared to a delayed intervention condition, to create positive changes in the following outcomes: *primary outcomes*: changes in confidence to engage in HIV prevention behaviors and changes in HIV knowledge and *secondary outcomes*: changes in sexual behavior and change in HIV testing behavior over the 12-month follow-up period. Descriptive statistics of the behavioral, psychosocial, and demographic characteristics of the participants will described for all and by group (intervention and control). These will be compared between treatment groups using t-tests or Wilcoxon rank sum tests for continuous variables and chi-square tests for categorical variables. In addition, we will use the general framework of generalized linear mixed models (GLMM) to test for intervention effects over time. Note that some of our outcomes are binary, some count and some continuous traits and thus need to be treated differently. The general form of the GLMM will be $(\mu_{ij})=\beta_{0i}+\beta_{\text{cov}}\ Covariates_{ij}+\ \beta_{\text{Time}}\text{Time}_j+\ \beta_{\text{Intervention}}\text{Group}_i+\ \beta_{\text{Group x Time}i}$, where μ_{ij} is the mean response corresponding to subject i at Time j (baseline and 4 follow-ups) with the appropriate link function (identity for continuous outcome, logit for binary outcome and natural log for count outcomes); Group =1 if the i-th subject is in the intervention group and 0 if the i-th subject is in the attention-control group. The interaction coefficients $\beta_{\text{Group x Time}}$ are of interest here, measuring the difference in the rate of change in outcomes across the two treatment groups over time in the primary and secondary outcomes. Models will be compared according to information criterion like AIC, BIC. For some binary outcomes like HIV testing, we will perform an aggregate analysis after collapsing across the repeated measures using simple logistic regression comparing whether the probability of having tested at least once over the entire FU period is different across treatment groups, after adjusting for baseline values. To ensure robustness, we will also apply an exchangeable working correlation structure to its corresponding generalized estimating equation (GEE) model.

Reporting Conventions

P-values ≥0.001 will be reported to 3 decimal places; p-values less than 0.001 will be reported as "<0.001". The mean, standard deviation, and any other statistics other than quantiles, will be reported to one decimal place greater than the original data. Quantiles, such as median, or minimum and maximum will use the same number of decimal places as the original data. Estimated parameters, not on the same scale as raw observations (e.g., regression coefficients) will be reported to 3 significant figures.