STATISTICAL ANALYSIS PLAN

RANDOMIZED CONTROLLED TRIAL OF AN LGBTQ-AFFIRMATIVE TREATMENT FOR YOUNG ADULT SEXUAL MINORITY MEN'S MENTAL AND SEXUAL HEALTH

STATISTICAL ANALYSIS PLAN

November 15, 2020

The SAP was prepared by the trial statisticians:

Denise Esserman
Jesse Reynolds
Kaitlin Maciejewski
Kirsty Clark
Richard Bränström

Revisions to the SAP were completed before data were locked.

STATISTICAL ANALYSIS PLAN

TABLE OF CONTENTS

STATISTICAL ANALYSIS PLAN

STUDY SUMMARY (revised 9/14/2020)

Title	A unified intervention for young gay and bisexual men's minority stress, mental health, and HIV risk (ESTEEM)
Study Design	The design is a three-arm, parallel-assignment randomized controlled trial design to test the efficacy of ESTEEM (Effective Skills to Empower Effective Men) with two comparison conditions—LGBTQ (lesbian, gay, bisexual, transgender, queer)-affirmative community mental health treatment (CMHT) and brief voluntary HIV counseling and testing (VCT) only. The unit of randomization is the individual.
Study Duration	5 years
Trial Sites	2 trial sites: New York City (Yale School of Public Health, Yale University) and Miami, Florida (University of Miami)
Objective	Conduct a randomized controlled trial to test the efficacy of a 10-session skills-building intervention designed to reduce young gay and bisexual men's co-occurring health risks by addressing the underlying cognitive, affective, and behavioral pathways through which minority stress impairs health.
Number of Subjects	The original target sample size was 250 participants enrolled from the two study sites [New York: 150 (60 ESTEEM, 60 CMHT, 30 VCT) and Miami: 100 (40 ESTEEM, 40 CMHT, 20 VCT)]. In our pilot study, we saw a 60% reduction in primary outcome at 6 months in the ESTEEM arm. We used these estimates to inform our 8-month endpoint. To achieve at least 90% power at a 5% type I error rate, accounting for an R-square of 0.1 between treatment arm and site, we expected needing 80 individuals in the ESTEEM and CMHT arms, and 40 individuals in the VCT-only arm. Although we planned to take steps to increase our retention rate from our pilot study, we conservatively estimated the retention at 8 months to be 80%, thus increasing the total sample size to a target sample size of 250 The target sample size was revised on 2/28/18 to correct an imbalance in randomization; recruitment ended after 44 months on DATE with a total of 254 participants enrolled.
Main Inclusion Criteria	(1) aged 18–35, (2) identify as a gay or bisexual man, (3) HIV-negative status confirmed through in-office testing, (4) diagnosis of *any Diagnostic and Statistical Manual (DSM-5)* depressive, anxiety, or trauma- and stressor-related disorder; (5) HIV sexual risk (≥1 act in past-90-day of condomless anal sex with a male partner of unknown status or HIV+ status, unless with a HIV+ primary/main partner with known undetectable viral load); (6) not currently adherent to PrEP (pre-exposure prophylaxis) (defined as taking 4 or more days per week) (7) NYC or Miami residential stability and planned availability for 12 months; and (8) provision of informed consent.

STATISTICAL ANALYSIS PLAN

Intervention	ESTEEM is a 10-session intervention based on the Unified Protocol, an individually-delivered cognitive behavioral therapy (CBT) intervention with efficacy for reducing stress-sensitive mental health disorders (e.g., depression, anxiety) by enhancing emotion regulation skills; reducing avoidance patterns; and improving motivation and self-efficacy for behavior change. The Unified Protocol employs modules for motivation enhancement, interoceptive and situational exposure, cognitive restructuring, mindfulness, and self-monitoring techniques. Through an extensive adaptation process, we adapted the Unified Protocol to enhance young gay and bisexual men's stigma coping by reducing minority stress processes.
Duration of Intervention	10 therapy sessions in 4 months.
Primary Outcome	The primary outcome is condomless anal sex in the absence of either PrEP or known undetectable viral load of HIV+ primary partners, measured with the Time-Line Follow-Back (TLFB), a semi-structured interview. The TLFB yields past-90-day incidence of HIV risk behavior: condomless anal sex, sex while using drugs or alcohol, number of sexual partners, and preceding-week PrEP use (i.e., coverage defined as 4+ doses per week). TLFB interviewers are masked to study arm. The primary outcome is assessed at 8-month follow-up.
Primary Analysis	We will use a fixed sequence procedure (gatekeeper strategy) to control for multiple testing of the primary comparisons: reduction in condomless anal sex in ESTEEM vs. VCT-only and ESTEEM vs. CMHT. We will conduct the comparison of ESTEEM versus VCT-only at the 0.05 level. If we find statistical significance between ESTEEM vs. VCT-only, then we will test the second comparison ESTEEM vs. CMHT at 0.05. To make use of all data collected, we will analyze the condomless anal sex outcome using a generalized linear mixed model with a logit link using a contrast to test the comparison at the primary time point of 8 months adjusting for site.
Secondary Outcomes	Depression severity, anxiety, drug abuse, alcohol abuse, mental health diagnoses, suicidality, minority stress and universal stress processes

STATISTICAL ANALYSIS PLAN

1. BACKGROUND

Young gay and bisexual men (YGBM) are at disproportionate risk of depression, anxiety, and substance use problems, which synergistically fuels their increasing risk of HIV infection. Among young men, approximately 93% of all diagnosed HIV infections are male-to-male. Mental health and substance use disparities drive YGBM's HIV- risk behavior. Despite HIV risk being influenced by mental health disparities for YGBM, no evidence based mental health intervention specifically tailored to this population exists. YGBM health interventions currently use a one-problem/one-treatment approach. Some interventions promote condom use, some encourage PrEP initiation and maintenance, and others reduce substance use. These treatments show moderate efficacy. None currently seeks to reduce mental health problems (e.g., depression, anxiety) among at-risk YGBM. A unified, transdiagnostic approach that addresses the pathways that unite these conditions may increase effectiveness, reduce cost, and provide a streamlined treatment experience for the most vulnerable YGBM, who are unlikely to seek multiple treatments for multiple health concerns.

ESTEEM (Effective Skills to Empower Effective Men) is a 10-session skills-building intervention designed to reduce YGBM co-occurring health risks by reducing the underlying cognitive, affective, and behavioral pathways through which minority stress impairs YGBM's health. ESTEEM is based on the Unified Protocol, a cognitive-behavioral therapy (CBT) approach with efficacy across mental health and risk behaviors. The Unified Protocol changes underlying stress pathways using motivational interviewing, emotional and situational exposure, cognitive restructuring, mindfulness, and self-monitoring exercises. To create ESTEEM, through an NIMH R34 award, our team adapted the Unified Protocol by conducting interviews with 21 YGBM-expert mental health providers and 20 depressed, anxious YGBM at high risk for HIV infection. These stakeholders helped our team infuse the Unified Protocol with minority stress coping content.34 ESTEEM aims to normalize the adverse impact of minority stress, reduce internalized homophobia and rejection schemas, reduce YGBM's unhealthy avoidance tendencies (e.g., substance use during sex, condom use non-assertion), and validate YGBM's unique strengths. In a preliminary trial (n=63), ESTEEM significantly reduced YGBM's spectrum of interrelated health threats, making it the first evidence-based intervention to simultaneously improve mental health, substance use, and sexual health outcomes among YGBM. However, important questions remain in order to validate the efficacy and potential cost-effectiveness of ESTEEM.

2. AIMS

1. **Aim 1:** Test the efficacy of ESTEEM against (1) community mental health treatment (CMHT) and (2) HIV/STD voluntary counseling and testing (VCT).

2. **Aim 2:** Determine whether ESTEEM works through its hypothesized cognitive, affective, and behavioral minority stress processes. 4-, 8-, and 12-month follow-ups will determine if improvements in minority stress processes precede and statistically mediate outcome improvements.

3. STUDY DESIGN

The design is a three-arm, parallel-assignment randomized controlled trial design to test the efficacy of ESTEEM with two comparison conditions—LGBTQ-affirmative community mental health treatment (CMHT) and brief voluntary HIV counseling and testing (VCT) only. The unit of randomization is the individual. The primary outcome is condomless anal sex in the absence of either PrEP or known undetectable viral load of HIV+ primary partners, measured with the Time-Line Follow-Back (TLFB), a semi-structured interview. The original target sample size was 250 participants from 2 study sites (NYC and Miami) detect 60% reduction in primary outcome at 8 months in the ESTEEM arm with 90% power. The target sample size was revised on 2/28/18 to correct an imbalance in randomization; study participation ended after 44 months on 6/24/20 with a total of 254 participants enrolled.

4. OUTCOMES

The primary and secondary outcome measures are summarized in Table 1.

Table 1. Primary and Secondary Outcome Measures

Domain	Measure (1˚,2˚, 3˚)	Source, Frequency, and Sample
HIV-risk behavior	Condomless anal sex (1˚; 2˚)	Interviewer-administered assessment at-home or in-office every 4 months. Condomless anal sex in the absence of either PrEP or known undetectable viral load of HIV+ primary partners, measured with the TLFB, a semi-structured interview. The TLFB yields past-90-day incidence of HIV risk behavior: condomless anal sex, sex while using drugs or alcohol, number of sexual partners, and preceding-week PrEP use (i.e., coverage defined as 4+ doses per week). Measured as any (binary) and number (count).
	Safe sex efficacy (2˚)	Safer Sex Self-Efficacy Scale; self-report every 4 months.
	Decisional balance (2˚)	Decisional Balance Scale; self-report every 4 months.
	PrEP uptake (2˚)	Interviewer-administered assessment every 4 months
	STI risk (3˚)	STI test results (gonorrhea; chlamydia) at 12 months
Mental health and substance use	Depression (2˚)	HAMD (interviewer-administered), ODSIS (self-report), BSI (self-report) every 4 months
	Anxiety (2˚)	BAI, OASIS, BSI, SIAS; self-report every 4 months
	Drug Abuse (2˚)	SIP-D; self-report every 4 months
	Alcohol Abuse (2˚)	AUDIT;cself-report every 4 months
	Mental Health Diagnoses (2˚)	MINI; interviewer-administered assessment every 4 months
	Suicidality (2˚)	SIDAS; self-report every 4 months
Minority Stress	Rejection Sensitivity (2˚)	RS; self-report every 4 months
	Internalized Homophobia (2˚)	IHS; self-report every 4 months
	Internalized Homophobia (2˚)	IAT; computer-administered every 4 months

	Sexual Orientation Concealment (2˚)	SOCS; self-report every 4 months
Universal Stress processes	Emotional Regulation (2˚)	DERS; self-report every 4 months
	Rumination (2˚)	RRS; self-report every 4 months
	Assertiveness (2˚)	RAS; self-report every 4 months
	Social Support (2˚)	MSPSS; self-report every 4 months
	Impulsivity (2˚)	Go/No-Go; computer-administered every 4 months
	Cognitive Bias (2˚)	SRET; computer-administered every 4 months

TLFB: Time-Line Follow-Back; MINI: Mini-International Neuropsychiatric Interview; HAM-D: Hamilton Rating Scale for Depression; BAI: SIP-D:; AUDIT ODSIS: Overall Depression Severity and Impairment Scale; OASIS: Overall Anxiety Severity and Impairment Scale; BSI: Brief Severity Inventory; DERS: Difficulties in Emotion Regulation Scale; RRS: Rumination Response Scale; MSPSS: Multidimensional Scale of Perceived Social Support; RAS: Rathus Assertive Schedule; RS: Rejection Sensitivity; IHS: Internalized Homophobia; SOCS: Sexual Orientation Concealment; IAT: Implicit Associations Task; SRET: Self-Referential Encoding Task; SIDAS: Suicidal Ideation Attributes Scale; SIAS: Social Interaction Anxiety Scale

4.1 Primary Outcome

The primary outcome is condomless anal sex at 8-month time point (binary: Yes/No) in the absence of either PrEP or known undetectable viral load of HIV+ primary partners, measured with the Time-Line Follow-Back (TLFB), a semi-structured interview. The TLFB will yield past-90-day incidence of HIV risk behavior: condomless anal sex, sex while using drugs or alcohol, number of sexual partners, and preceding-week PrEP use (i.e., coverage defined as 4+ doses per week). The primary outcome is measured every 4 months (4, 8, 12 months).

4.2 Secondary Outcomes

Condomless anal sex at 12 months (binary); Condomless anal sex at 8 and 12 months (count); Depression severity, as measured using the HAMD depression severity score, ODSIS, and BSI at 8 and 12 month time points; Anxiety, as measured by the BAI, OASIS, BSI, and SIAS at 8 and 12 month time points; Drug abuse, as measured by the SIP-D at 8 and 12 month time points; Alcohol abuse, as measured using by the AUDIT at 8 and 12 month time points; Mental health, as defined by odds of no longer having at least one *DSM-5* diagnosis (e.g., depression, anxiety) that was present at baseline still be present at 8 and 12 month time points based on the MINI (binary); Suicidality as measured by SIDAS at 8 and 12 month time points (continuous); Minority stress as captured by rejection sensitivity measured using RS, internalized homophobia measured using IHS and IAT; and sexual orientation concealment measured using SOCS at 4 and 8 month time points (continuous); Universal stress processes as captured by emotional regulation as measured by DERS, rumination as defined by RRS, assertiveness as measured by RAS, social support as measured by MSPSS, cognitive bias as measured by SRET, and impulsivity measured by Go/No-Go at 4 and 8 month time points (continuous).

4.3 Tertiary Outcomes

STI risk as defined by a diagnosis of gonorrhea; chlamydia) at 12 months; mental health mediators (as defined for the secondary outcomes).

5. RANDOMIZATION

5.1 Method of Randomization

Randomization occurred on the participant level. Participants were randomly assigned to receive one of the following three conditions: ESTEEM, CMHT, or VCT-only (N = 250 across both sites). We used a 2:2:1 randomization scheme such that for every two participants randomized to ESTEEM, two participants were randomized to CMHT and one participant was randomized to VCT-only. Randomization happened through a computer-generated program stratified by site, such that 150 were randomized at the NYC site (60 ESTEEM, 60 CMHT, 30 VCT-only) and 100 were randomized at the Miami site (40 ESTEEM, 40 CMHT, 20 VCT-only). On 2/28/18, we discovered a programming error in the randomization and participants were not being randomized 2:2:1. The randomization was corrected to ensure the final distribution was 2:2:1 and the sample size was increased to 254 to accommodate.

5.2 Allocation Concealment

An important consideration is allocation concealment to control for selection bias. Potential approaches include 1) randomizing after enrollment, 2) masking the recruiters, 3) standardizing the enrollment process with adequate training of screeners and recruiters, and 4) covariate adjustment in the analysis. We used approaches 1 through 3, and will employ approach 4, if necessary. Specifically, Qualtrics was initially used to randomize participants but was found to be doing so in error, using a 1:1:1 allocation, rather than the proposed 2:2:1. In February 2018, the following update was made**:** Team members agreed to a solution whereby the study changed the randomization scheme to correct for the previous 1:1:1 randomization and slow the rate of participant randomization to VCT. This was done through the use of a randomization strategy that would increase the probability of participants being assigned to ESTEEM and CMHT, and that by the end of the study would yield a 2:2:1 ratio. Based on the number of participants that were randomized from each site and their current allocation to treatment arms, it was determined that a 3:3:1 ratio of randomization should be used at each site to ensure the originally proposed balance of 2:2:1 be met by study's end. Further, it was recommended that participants be randomized in blocks at each site of 7 (3:3:1). This provision made sure that the each site retained the correction ratio consistently through the end of recruitment. At the time, to meet recruitment goals, 54 participants were remaining to be enrolled at the NYC site and 82 participants were remaining to be enrolled at the Miami site. As a precaution, the statistician recommended that each site complete a recruitment block of 7 regardless of targeted enrollment numbers. Using the remaining participants and the original target enrollment, it was suggested that NYC enroll *at least* 56 more participants (8*7) and Miami enroll *at least* an additional 84 participants (12*7). Using the SAS procedure 'Proc Plan,' 2 lists were generated per site using the block randomization method described. Based on the suggested allocation per site, a total of 140 participants were randomized to complete the study to provide a 2:2:1 balance. Envelopes containing the randomization codes were created numbered 1 to 140 per site. These envelopes were provided to each site with a tracking sheet (site list). Each envelope contained an assignment that matches each site list. When each participant was deemed eligible, the next envelope (in sequential order) was opened and treatment assigned. The

assignment was logged in the tracking sheet at the site and that study statistician was notified of assignment. An identical tracking sheet was used by the statistician at Yale. To track the balance of the adjusted randomization moving forward, periodic 'check ins' between the sites and study statistician monitored the progress and these updates in enrollment numbers and study arm balance was provided to the DSMB.

Recruiters and assessors/interviewers were masked to treatment assignment up until the point in which they were enrolled.

6. SAMPLE SIZE

6.1 Preliminary Data

Sample size estimates were informed by preliminary data from a pilot of the ESTEEM study (Pachankis et al. 2015), in which we saw a 60% reduction in the primary outcome at 6 months in the ESTEEM arm compared to waitlist control.

6.2 Sample Size Determination for the Primary Outcome

Sample size calculations were carried out using PASS 12 for logistic regression.

Our primary goal is to demonstrate a greater reduction at 8 months in condomless anal sex in the absence of either PrEP or known undetectable viral load of HIV+ primary partners, in the ESTEEM arm versus the CMHT and VCT-only arms. In our pilot study, we saw a 60% reduction in condomless anal sex at 6 months in the ESTEEM arm. We used these estimates to inform our 8-month endpoint.

Based on previous studies of VCT and the fact that CMHT does not specifically focus on condomless anal sex, we expect that these arms will yield lower reductions compared to ESTEEM, but a slightly larger reduction in CMHT (20%) compared to VCT-only (15%). To achieve at least 90% power at a 5% type I error rate, accounting for an R-square of 0.1 between treatment arm and the covariates (e.g., site, race/ethnicity), we will need 80 individuals in the ESTEEM and CMHT arms, and 40 individuals in the VCT-only arm. Although we plan to take steps to increase our retention rate from our pilot study, we conservatively estimate the retention at 8 months to be 80%. Therefore, we planned to randomize 100 ESTEEM,100 CMHT, and 50 VCT-only young gay and bisexual men. The original target sample size was 250 participants enrolled from the two study sites [New York: 150 (60 ESTEEM, 60 CMHT, 30 VCT) and Miami: 100 (40 ESTEEM, 40 CMHT, 20 VCT)]. This was increased to 254 on 2/28/18 to correct an imbalance in the 2:2:1 randomization. The final sample size was 254.

6.3 Power for Secondary Outcomes

For secondary outcomes (e.g., mental health, substance use), we will have 80% power to detect an effect size of 0.55 and 0.70 for ESTEEM vs. CMHT and ESTEEM vs. VCT-only, respectively, at a type I error rate of 0.01 (conservative due to multiple testing). These effect sizes are smaller than those found in the pilot (Pachankis et. al. 2015).

6.4 Power for Tertiary Outcomes

Power analyses were not conducted to inform tertiary outcomes.

7 INTERIM MONITORING

7.1 Overview

Interim monitoring focused on patient accrual, baseline comparability of treatment groups, protocol adherence, loss to follow-up, data completeness and quality, safety, efficacy, and futility. The summary of the types of tables, listing and figures (TLFs) generated for semi-annual open and closed DSMB reports is presented in Appendix A.

7.2 Safety

Details of the safety and adverse event monitoring plan were provided in study protocol. The principal investigator was responsible for monitoring the data and assuring protocol compliance. Unanticipated problems involving risks to subjects or others, including adverse events, were followed by a written report within five calendar days of the principal investigator becoming aware of the event to the institutional review board (IRB). The principal investigator apprised fellow investigators and study personnel of all unanticipated problems and adverse events that occurred during the conduct of this research project. The protocol's data safety monitoring board (DSMB) was informed of serious or unanticipated adverse events. Presented to the DSMB at each annual interim meeting were the following summaries by time point: HIV status, incidence of chlamydia and gonorrhea, TLFB reports of risky sexual behavior (total sex acts, total sex acts under the influence and condomless anal sex acts), and summaries of depression and suicidality scores (HAMD and SIDAS respectively). These reports were presented across treatment arms in aggregate in an 'open meeting' with the study team present, followed by an unblinded presentation of these data by treatment arm to the DSMB by the unblinded statistician.

8 ANALYTIC PLAN

8.1 Overview

The analysis of the primary and secondary outcomes will be according to the principle of intent-to-treat, i.e., participants will be analyzed according to their original treatment assignment regardless of adherence to protocol. All analyses will include the participant as the unit of analysis. SAS 9.4, SPSS 26.0, MPlus 8.4, and the latest version of R (currently 4.0.2) software will be used for all analyses.

8.2 Comparability of Treatment Groups

To determine randomization effectiveness, differences in baseline demographic characteristics will assessed between ESTEEM, CMHT, and VCT-only groups using appropriate graphical and statistical methods including summary statistics. The randomization is designed to produce balance on important covariates. If we find imbalance on key covariates, we will consider sensitivity analyses that adjust for those covariates. Study site will be included as a covariate in all analyses because randomization was stratified by study site.

8.3 Analysis of Primary Outcome: HIV-Risk Behavior (Condomless Anal Sex)

8.3.1 Primary Analysis of Primary Outcome: Condomless Anal Sex

Analysis	Primary analysis of primary outcome: condomless anal sex in the absence of either PrEP or known undetectable viral load of HIV+ primary partners, measured with the TLFB, a semi-structured interview.
Analysis population	All enrolled participants
Endpoint	8-month as reported on TLFB
Unit of analysis	Participant
Method of analysis*	Generalized linear mixed model with a log (log-binomial) link using a contrast to test the comparison at the primary time point of 8 months adjusting for site and race/ethnicity (Wacholder 1986). We will conduct the comparison of ESTEEM versus VCT at 0.05 level; only if this test is statistically significant will we test ESTEEM versus CMHT at 0.05.
Handling of missing data	MAR assumption
Adjustment covariates	Study site (NYC or Miami).
Type I error	5% (2-sided)
Control of type I error	We will use a gatekeeping strategy. If the first comparison is significant at 0.05, then we will conduct the second comparison at 0.05.
Treatment effect estimate	Risk ratio with 95% confidence limits

*If the log-binomial approach fails to converge, we will use a modified log-Poisson model (Zou 2004).

Sensitivity analysis: This will only be conducted if baseline imbalance of covariates is found (see Section 7.2). Adjustment for baseline covariates not balanced at randomization.

8.3.2 Supportive Analysis of the Primary Outcome: Number of Condomless Anal Sex Acts

Analysis	Supportive analysis of primary outcome: number of condomless anal sex acts in the absence of either PrEP or known undetectable viral load of HIV+ primary partners, measured with the TLFB, a semi-structured interview.
Analysis population	All enrolled participants
Endpoint	8-month as reported on TLFB
Unit of analysis	Participant
Method of analysis*	Generalized linear mixed model with a log link (Poisson model) using a contrast to test the comparison at the primary time point of 8 months adjusting for site and baseline number of sex acts. We will conduct the comparison of ESTEEM versus VCT at 0.05 level; only if this test is statistically significant will we test ESTEEM versus CMHT at 0.05.
Handling of missing data	MAR assumption
Adjustment covariates	A priori covariates include: study site (NYC or Miami); baseline number of sex acts
Type I error	5% (2-sided)
Control of type I error	We will use a gatekeeping strategy. If the first comparison is significant at 0.05, then we will conduct the second comparison at 0.05.
Treatment effect estimate	Risk ratio with 95% confidence limits

*We will assess for overdispersion. If there is a large amount of overdispersion, we will correct the Poisson model to take this into account. We will also assess for zero-inflation. If there are excess zeros, we will use a zero-inflated Poisson model.

8.4 Analysis of Secondary Outcomes

Secondary outcomes of interest: Condomless anal sex at 12 months (binary); Condomless anal sex at 8 and 12 months (count); Depression severity, as measured using the HAMD depression severity score, ODSIS, and BSI at 8 and 12 month time points; Anxiety, as measured by the BAI, OASIS, BSI, and SIAS at 8 and 12 month time points; Drug abuse, as measured by the

SIP-D at 8 and 12 month time points; Alcohol abuse, as measured using by the AUDIT at 8 and 12 month time points; Mental health, as defined by odds of no longer having at least one *DSM-5* diagnosis (e.g., depression, anxiety) that was present at baseline still be present at 8 and 12 month time points based on the MINI (binary); Suicidality as measured by SIDAS at 8 and 12 month time points (continuous); Minority stress as captured by rejection sensitivity measured using RS, internalized homophobia measured using IHS and IAT; and sexual orientation concealment measured using SOCS at 4 and 8 month time points (continuous); Universal stress processes as captured by emotional regulation as measured by DERS, rumination as defined by RRS, assertiveness as measured by RAS, social support as measured by MSPSS, cognitive bias as measured by SRET, and impulsivity as measured by Go/No-Go at 4 and 8 month time points (continuous). Each outcome will be analyzed separately as described below.

8.4.1 Condomless anal sex at 12 months (binary and count)

These outcomes will follow the same methodology outlined above for the primary outcome.

8.4.2 Mental Health and Substance Use

Mental health and substance use included: depression severity (measured by HAMD, ODSIS, BSI); Anxiety (measured by BAI, OASIS, BSI, SIAS); Drug abuse (measured by SIP-D); Alcohol abuse (measured by AUDIT); Mental health (measured using MINI); and suicidality (measured by SIDAS).

8.4.2.1 Continuous Measures

Analysis	Mental Health and Substance Use: depression severity; anxiety; drug abuse; alcohol abuse; suicidality
Analysis population	All participants
Endpoint	Scale values at 8 and 12 months of follow-up
Unit of analysis	Participant
Method of analysis	Linear mixed model assuming missing at random (MAR)
Handling of missing data	MAR assumption
Adjustment covariates	A priori covariates include: study site (NYC or Miami); baseline score
Type I error	5% (2-sided)
Control of type I error	To control the false discovery rate, we will use the Benjamini and

	Hochberg (1988) method.
Treatment effect estimate	Marginal least square mean (LSM) with 95% confidence limits if there is no treatment by time interaction ($p<0.10$); otherwise, LSM will be summarized for each treatment and timepoint.

Sensitivity analyses will be conducted to investigate the MAR assumption, such as methods that model jointly the missingness and outcome distributions (National Research Council, 2010).

8.4.2.2 Binary Measure

Use of the MINI to define having at least one diagnosis present at baseline, no longer present at follow-up. We will use method described above for the primary outcome and assess at 8 and 12 months.

8.4.3 Minority Stress- Mediation analysis

We plan to model minority stress mediators (rejection sensitivity, internalized homophobia, sexual orientation concealment) as a latent variable at the 4- and 8-month time points. We will then use these constructs in latent variable mediation models using structural equation modeling to examine whether 4-month minority stress mediates 8-month sexual risk and mental health outcomes; and whether 8-month minority stress mediates 12-month sexual risk and mental health outcomes. The mediation models will include test of exposure-mediation interactions (Valeri and VanderWeele 2013; Muthan and Asparoutov 2015), to consider potential confounding of the mediation-outcome relationship and allow for causal interpretation. In case of support for a mediating effect of minority stress in the above latent variable mediation model, we will apply a sequential approach in our analyses of mediation to estimate both the joint effect through all mediations, as well as, give consideration to how much of the combined effect is attributed to each mediator (Vansteelandt and Daniel 2017; VanderWeele and Vansteelandt 2014).

8.4.4 Universal Mediators- Mediation analysis

We plan to model universal mediators (DERS, RRS, RAS, MSPSS, cognitive bias, impulsivity) as a latent variable at the 4- and 8-month time points. We will then use these constructs in latent variable mediation models using structural equation modeling to examine whether 4-month universal mediators mediate 8-month sexual risk and mental health outcomes; and whether 8-month universal mediators mediates 12-month sexual risk and mental health outcomes. The mediation models will include test of exposure-mediation interactions (Valeri and VanderWeele 2013; Muthan and Asparoutov 2015) to consider potential confounding of the mediation-outcome relationship and allow for causal interpretation. In case of support for a mediating effect of minority stress in the above latent variable mediation model, we will apply a sequential approach in our analyses of mediation to estimate both the joint effect through all mediations, as well as, give consideration to how much of the combined effect is attributed to each mediator (Vansteelandt and Daniel 2017; VanderWeele and Vansteelandt 2014).

Mediation analysis

8.5 Analysis of Tertiary Outcomes

8.5.1. STI Risk

STI risk (gonorrhea; chlamydia) measured at 12 months (yes/no). We will use the same analysis techniques as proposed for the primary binary outcome (see Section 8.3.1). If there are few events and it is not possible to adjust for site, we will conduct a chi-square test of association.

8.5.2 Mental health mediators

We will follow a similar analysis plan as proposed in 8.4.3 and 8.4.4 using the mental health variables as the mediators and determining the mediating relationship on HIV risk outcomes using both latent variable mediation models and mediation analysis with multiple mediators.

8.6 Analysis of Safety

Safety involved tabulating HIV status, incidence of chlamydia and gonorrhea, TLFB reports of risky sexual behavior (total sex acts, total sex acts under the influence and condomless anal sex acts), and summaries of depression and suicidality scores (HAMD and SIDAS respectively).

REFERENCES

Hochberg Y (1988). A sharper Bonferroni procedure for multiple tests of significance. Biometrika. 75:800-802

Muthén, B. and T. Asparouhov (2015). "Causal effects in mediation modeling: An introduction with applications to latent variables*." Structural Equation Modeling: A Multidisciplinary Journal* 22(1): 12-23.

National Research Council (2010). The prevention and treatment of missing data in clinical trials. National Academies Press.

Pachankis, J. E., Hatzenbuehler, M. L., Rendina, H. J., Safren, S. A., & Parsons, J. T. (2015). LGB-affirmative cognitive-behavioral therapy for young adult gay and bisexual men: A randomized controlled trial of a transdiagnostic minority stress approach. *Journal of consulting and clinical psychology*, *83*(5), 875.

Valeri, L. and T. J. VanderWeele (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychological methods* 18(2): 137.

Vansteelandt, S. and R. M. Daniel (2017). Interventional effects for mediation analysis with multiple mediators. *Epidemiology* (Cambridge, Mass.) 28(2): 258.

VanderWeele, T., & Vansteelandt, S. (2014). Mediation analysis with multiple mediators. *Epidemiologic methods*, *2*(1), 95-115.

Wacholder S. (1986). Binomial regression in GLIM: estimating risk ratios and risk differences. *Am J Epidemiol. Jan; 123(1):174-84.*

Zou G. (2004). A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol. Apr 1; 159(7):702-6.*