A Multisectoral Strategy to Address Persistent Drivers
of the HIV Epidemic in East Africa (SAPPHIRE)

Sustainable East Africa Research in Community Health (SEARCH) Consortium

Statistical Analysis Plan

Project: Effect of transport voucher and missed visit phone call on linkage to hypertension care in SEARCH SAPPHIRE

Laura B. Balzer, PhD*
Diane V. Havlir, MD**
Maya L. Petersen, PhD***
Matthew D. Hickey, MD**

for the SEARCH Collaboration

December 21, 2021

Version 1.0

*School of Public Health & Health Sciences, University of Massachusetts Amherst, Amherst, USA
**Division of HIV, ID, & Global Medicine, Department of Medicine, UCSF, San Francisco, USA
***School of Public Health, University of California Berkeley, Berkeley, USA

1. Overview

The SEARCH SAPPHIRE Hypertension Linkage study is an individual randomized controlled trial, designed to test the hypothesis that a transportation voucher and a follow-up phone call will improve linkage to clinic-based hypertension care following community-based screening. The trial takes place in three communities in Kenya and Uganda and began enrollment in June 2021. After community-based screening, individuals diagnosed with hypertension (HTN) were enrolled and given a clinic appointment at the nearest government-run clinic. Participants randomized to intervention (n=100) received a voucher to reimburse transport expenses, redeemable upon clinic linkage at any time following enrollment (even if initial linkage appointment was missed). Intervention participants who missed their initial linkage appointment also received a follow-up phone call to encourage linkage. Control participants (n=100) did not receive a transport voucher or any follow-up if they missed their initial linkage appointment. Upon linkage to care, all participants received integrated, patient-centered hypertension care, detailed elsewhere.[1–3] Treatment guidelines for use of antihypertensive medication were based on standard country guidelines. Additional details of the study procedures are available in the Study Protocol.

The primary objective of this study is to determine whether the intervention improved linkage to HTN care by 30 days. Secondary endpoints include linkage at or before the scheduled appointment, linkage by 3 months, HTN control at follow-up, and retention in hypertension care. We will also characterize changes in HTN severity and evaluate predictors of linkage, HTN control, and retention.

Throughout, the population of interest comprises non-pregnant adults, aged 25 years and older, who do not have a prior diagnosis of hypertension, and who have uncontrolled HTN at baseline as identified through community-based screening (defined as systolic blood pressure ≥140 mmHg or diastolic blood pressure ≥90 mmHg on all three baseline measures). Participants were excluded if they had a blood pressure ≥180/110mmHg and had symptoms of hypertensive emergency – these individuals were immediately transported to the nearest health facility for emergent treatment.

We will provide descriptions of participant flow through the study (i.e., a consort diagram), measurement coverage, and baseline characteristics (e.g., country, age, sex, marital status, occupation, monthly income, education, comorbidities, and baseline severity of hypertension), overall and by arm and sex. We will also report risk factors (N and %) for uncontrolled HTN at baseline, overall and by sex. Additionally, we will report and describe intervention fidelity and delivery of clinical hypertension care. Finally, we will report and describe barriers and facilitators of hypertension care engagement, as assessed through participant self-report at baseline and follow-up.

2.1 General approach for evaluating intervention effects

We will assess the intervention effect with targeted minimum loss-based estimation (TMLE), which provides precision and power gains over an unadjusted analysis by adjusting for baseline predictors of the outcome.[4–7] For a detailed review of TMLE and its relation to other effect estimators in randomized trials, we refer the reader to Colantuoni and Rosenblum.[8] For a recent demonstration of the improved precision offered by TMLE in randomized trials, we refer the reader to Balzer et al. and to Benkser et al..[9,10]

Here, we will use **TMLE with Adaptive Pre-specification**, a fully automated procedure to flexibly adjust for baseline outcome predictors, while maintaining Type-I error control.[11] Specifically, using 10-fold cross-validation, we will chose the optimal approach for estimating the expected outcome given the randomization arm and baseline covariates (a.k.a., the outcome regression) and for estimating the conditional probability of being randomized to the intervention given the baseline covariates (a.k.a., propensity score). Throughout, optimality is defined by using the squared influence curve for the TMLE as loss function. Thereby, we will select the combination of estimators (adjustment variables + approach) of the outcome regression and of the propensity score that minimizes the cross-validated risk estimate and, thus, the cross-validated variance estimate.

Our pre-specified, candidate adjustment variables are age, sex, hypertension severity ("grade 1" as 140-159/90-99 mmHg versus "grade 2 or higher" as ≥160/100 mmHg), site, and nothing (i.e., unadjusted). Our pre-specified, candidate estimators of the outcome regression are main terms, stepwise regression, stepwise regression with all possible pairwise interactions, LASSO, and the mean. Our pre-specified, candidate estimators of the propensity score are main terms, stepwise regression, LASSO, and the mean. In sensitivity analyses, we will also implement that unadjusted effect estimator as the contrasts of average outcomes by arm.

For all endpoints, primary estimates will be for the study sample and on the **relative scale**: $1/n \sum_{i=1}^{n} Y_i(1) \div 1/n \sum_{i=1}^{n} Y_i(0)$, where $Y_i(1)$ denotes the counterfactual outcome for participant *i* under the intervention and $Y_i(0)$ denotes the counterfactual outcome for participant *i* under the control. Secondary comparisons will be on the absolute scale.

For all endpoints, we will test the **null hypothesis** of no improvements in outcomes due to the intervention with a one-sided test at the 5% significance level. We will also report point estimates and 95% confidence intervals for each effect measure and the arm-specific average outcomes. Standard error estimation will be with the estimated influence curve, and statistical inference will follow from the Central Limit Theorem (i.e., using the standard normal distribution).[4]

For all endpoints, we will also examine the intervention effect within **subgroups** defined by sex, age group (<60 years vs. ≥60 years), country, site, baseline HTN severity, and HIV status. In subgroups with fewer than 41 participants, we will limit the candidate estimation approaches to main terms adjustment for a single covariate or the simple mean.

2.2 Evaluate the effect on linkage to hypertension care at 30 days

The primary endpoint is linkage to care by 30 days, evaluated with clinical records. Participants without clinical records indicating linkage to care will be assumed never to have linked. Sensitivity analyses will incorporate self-report, defined as having a clinical record of linkage by 90 days or self-reported linkage to any health facility within the past 3 months. These linkage endpoints are not subject to missingness.

As described above, we will use TMLE with Adaptive Pre-specification to compare linkage to care by 30 days between intervention and control arms and to test the null hypothesis that the intervention did not improve linkage to care by 30 days. Sensitivity analyses incorporating self-report and subgroup analyses will be implemented analogously.

We will also provide descriptive statistics of participants who linked to care by 30 days. Finally, we will generate a graphic of cumulative linkage over time by arm.

2.3 Evaluate the effect on linkage to scheduled post-screening appointment

To assess the effect of the transport voucher, we will again use TMLE to compare linkage to care by the scheduled appointment date between intervention and control arms. We will use the same candidate adjustment variables and approaches as the primary endpoint. This endpoint is also not subject to missingness. We will formally test the null hypothesis that the intervention did not improve linkage to care on or before the scheduled linkage appointment. We will also implement analogous subgroup analyses and provide descriptive statistics of participants who linked to care by their scheduled appointment date.

To elucidate the added value of the transport voucher, we will also compare the intervention effect on linking by 30 days (overall) to the intervention effect on linkage to the scheduled appointment. To do so, we will use the Delta Method to test the null hypothesis that the ratio of relative risks is 1 and that the difference of absolute effects is 0. Here, we will use a two-sided test at the 5% significance level.

2.4 Evaluate the effect on hypertension control at months 3, 6, and 12

Again using TMLE with Adaptive Pre-specification, we will assess the intervention effect on HTN control (all three blood pressure measures <140/90 mmHg) at months 3, 6 and 12 post-enrollment. We will formally test the null hypothesis that the intervention did not improve hypertension control at these timepoints. We will use the same candidate adjustment variables and approaches as the primary endpoint (linkage within 30 days).

In the primary analysis of hypertension control, participants who do not have their blood pressure measured within the endpoint date (e.g., in the window of 2.5-3.5 months post-enrollment for the 3-month endpoint) will be considered to have uncontrolled hypertension. In sensitivity analyses, we will control for incomplete ascertainment of blood pressure measures. These secondary analyses will use TMLE to estimate the intervention effect, while adjusting for differences in characteristics (e.g., arm, sex, age, severity, site) between persons with measurements and persons with missing measurement.

We will again implement subgroup analyses as well as examine control among participants who linked to HTN care within 30 days. We will also evaluate the

intervention impact on HTN severity, defined as participants with grade 2 or higher HTN at follow-up, and on the average of the second and third systolic blood pressure measures at follow-up.

2.5 Evaluate the effect on retention in hypertension care at 3, 6 and 12 months

We will define retention in care as not late for the most recent scheduled hypertension care appointment by 30 days or more. This endpoint is also not subject to missingness. Again, we will use the primary analytic approach of TMLE with Adaptive Pre-specification to assess the intervention effect on retention in hypertension care at 3, 6 and 12 months and to test the null hypothesis that the intervention did not improve retention in care at 3, 6 or 12 months, respectively. For this endpoint, the candidate adjustment variables will be expanded to include baseline HIV status. We will implement analogous subgroup analyses.

2.6 Assess individual-level predictors of linkage to care, control, and retention in care

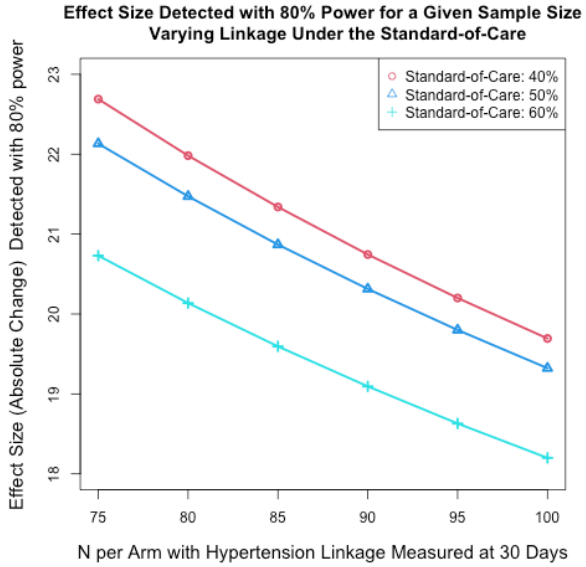
We will evaluate the following predictors of linkage to care within 30 days: sex, age, country, baseline HTN severity (grade), and baseline HIV status (assessed via testing and self-report). In the primary analysis, TMLE will be used to obtain variable importance measures, capturing the amount of information that a given predictor provides after adjusting for the other predictors. For each of the predictors, we will report the adjusted variables importance measures on the relative scale, treating each baseline predictor in turn as the “exposure” variable, and the rest as the adjustment set. In secondary analyses, we will calculate and report unadjusted (i.e., univariate) associations. We will conduct these analyses overall and stratified by randomized arm. Pooled analyses will include randomization arm as a predictor.

Analogous methods are used to evaluate the predictors of HTN control at 3, 6, and 12 months and retention in care at 3, 6 and 12 months.

Appendix: Power calculations

Sample size and power calculations were based of standard formulas for a two-sample test of proportions and done with *power.prop.test* in *R*.[12] We expect these calculations to be conservative, because of the precision gained through stratified randomization, through covariate adjustment during the analysis, and through our use of a one-sided hypothesis test.

We estimated 100 participants/arm would provide 80% power to detect at least a 19.3% absolute increase in linkage within 30 days from 50% under the control. As shown in the following Figure, even with 25% fewer participants enrolled (from 100 to 75 participants/arm) and lower or higher linkage under the standard-of-care, these calculations suggest we would be well-powered to detect at least a 22.7% absolute increase in linkage.



References

1. Kwarisiima D, Atukunda M, Owaraganise A, et al. Hypertension control in integrated HIV and chronic disease clinics in Uganda in the SEARCH study. *BMC Public Health*. 2019;19(1):511. doi:10.1186/s12889-019-6838-6

2. Kwarisiima D, Kamya M, Owaraganise A, et al. High rates of viral suppression in adults and children with high CD4+ counts using a streamlined ART delivery model in the SEARCH trial in rural Uganda and Kenya. *J Int AIDS Soc*. 2017;Jul 21(20).

3. Havlir DV, Balzer LB, Charlebois ED, et al. HIV Testing and Treatment with the Use of a Community Health Approach in Rural Africa. *N Engl J Med*. 2019;381(3):219-229.

4. van der Laan M, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer; 2011.

5. Moore KL, van der Laan MJ. Covariate Adjustment in Randomized Trials with Binary Outcomes: Targeted Maximum Likelihood Estimation. *Stat Med*. 2009;28(1):39-64. doi:10.1002/sim.3445

6. Rosenblum M, van der Laan MJ. Simple, Efficient Estimators of Treatment Effects in Randomized Trials Using Generalized Linear Models to Leverage Baseline Variables. *Int J Biostat*. 2010;6(1):Article 13. doi:10.2202/1557-4679.1138

7. van der Laan MJ, Rose S. *Targeted Learning in Data Science*. Springer; 2018.

8. Colantuoni E, Rosenblum M. Leveraging Prognostic Baseline Variables to Gain Precision in Randomized Trials. *Stat Med*. 2015;34(18):2602-2617.

9. Balzer LB, van der Laan M, Ayieko J, et al. Two-Stage TMLE to Reduce Bias and Improve Efficiency in Cluster Randomized Trials. *Biostatistics*. 2021;kxab043. https://doi.org/10.1093/biostatistics/kxab043

10. Benkeser D, Díaz I, Luedtke A, Segal J, Scharfstein D, Rosenblum M. Improving precision and power in randomized trials for COVID-19 treatments using covariate adjustment, for binary, ordinal, and time-to-event outcomes. *Biometrics*. 2021;n/a(n/a):1-15. doi:10.1111/biom.13377

11. Balzer L, van der Laan MJ, Petersen M, SEARCH Collaboration. Adaptive Pre-specification in Randomized Trials With and Without Pair-Matching. *Stat Med*. 2016;35(10):4528-4545. doi:10.1002/sim.7023

12. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2020. http://www.R-project.org

Sustainable East Africa Research in Community Health (SEARCH) Collaboration

Statistical Analysis Plan for Dynamic Choice Care Intervention for Mobile Persons with HIV in Phase A of SEARCH-Sapphire

Laura B. Balzer, PhD[1]
and the SEARCH Study Team[2]

May 19, 2022

v1.0

[1]University of California, Berkeley
[2]www.searchendaids.com

Table of Contents:

1. Study Overview

In Phase A of SEARCH-Sapphire (NCT04810650), we are conducting an individually randomized controlled trial to evaluate the effect of dynamic choice, patient-centered care intervention for mobile persons with HIV (PWH) in rural Kenya and Uganda. Details of the trial design and procedures can be found in the corresponding Study Protocol. Analyses plans for qualitative outcomes and cost-effectiveness outcomes are available elsewhere. Power calculations are given in the Appendix.

In brief, from April through July 2021, we enrolled 201 mobile PWH who were at risk of viral non-suppression. These participants were randomized to the intervention or the standard-of-care using a stratified random block design with stratification factors of country and sex and with random block sizes of 2 and 4. The randomization list was generated by an independent researcher.

The intervention includes choice of the following components to address barriers to HIV care among mobile PWH:

- Travel pack, including emergency antiretroviral therapy (ART) supply, discrete ART packaging, and a travel checklist
- Hotline access to a mobility coordinator
- Offsite and longer ART refills
- Facilitated access to ART and HIV care outside of the community

The intervention is delivered by a mobility coordinator, stationed at each clinic, and follow-up is over 48 weeks.

The primary objective is to evaluate if the intervention improved viral suppression (<400 copies/mL) among mobile PWH at week-48. Secondary endpoints, compared

between randomized arms, include retention in HIV care and ART possession. Additionally, within the intervention arm, we will report coverage and uptake of the intervention components at both scheduled and unscheduled visits.

2. Population and Characteristics

The population of interest is PWH who are aged 15+ years, enrolled or new to HIV care, at risk of viral non-suppression (HIV RNA>400 copies/mL in the past year or 2+ missed HIV care visits in the past year), and mobile (2+ weeks out of the community in the past year).

To characterize measurement of this population, we will provide a participant flow diagram (i.e., a CONSORT diagram). Overall and stratified by trial arm and further by sex, we will summarize the baseline characteristics, including sex, age, country, marital status, occupation, education level, alcohol use (any use in past 3 months), mobility metrics (nights away in past 3 months and in past 12 months), impact of mobility on HIV care (missed visits and missed ART doses), baseline ART status and regimen, baseline viral suppression status, and trial enrollment criteria. We will categorize age as "younger" if aged 15-30 years and define persons to be "highly mobile" if they report spending >14 nights away from the community in the past 3 months.

3. Endpoint Measurement and Definition

The **primary endpoint** is HIV viral suppression (HIV RNA<400 copies/mL) at 48-weeks. The primary analytic population consists of all living study participants who have not transferred HIV care to another health facility (as documented in clinical records). In other words, the primary analysis will exclude persons who died, withdrew, or transferred care, but will, otherwise, include persons regardless of their level of mobility or history of movement in/out of the community. In the primary analysis, missing endpoint viral loads will be treated as failures (i.e., unsuppressed). In pre-specified sensitivity analyses, described below, we will assess the robustness of these analytic choices.

Based on review of Ministry of Health medical and pharmacy records, we will also examine the following **secondary endpoints**:

- Retention in care: the proportion of follow-up time where the participant is engaged in clinical care
- ART possession: the proportion of follow-up time where the participant has a full regimen of ART

We will calculate total follow-up time for each participant as the number of days from their trial enrollment to their week-48 viral load measure. (For persons without a week-48 viral load measure, we will use the close the endpoint ascertainment window.) Out-of-care time will start 14 days after a missed visit and end at re-engagement in care.

4. Evaluation of the SEARCH Intervention Effect

We will assess the intervention effect with targeted minimum loss-based estimation (TMLE), which improves precision and power by adaptively adjusting for baseline outcome predictors.[1–5] Here, we will use **TMLE with Adaptive Pre-specification** to flexibly control for baseline covariates, while maintaining Type-I error control and accounting for the randomization scheme.[6–8] Using 10-fold cross-validation, we will chose the optimal approach for estimating the outcome regression (i.e., the expected outcome given the randomization arm and adjustment covariates) and the known propensity score (i.e., the conditional probability of being randomized to the intervention given the adjustment covariates). Specifically, we will select the combination of estimators (adjustment variables + approach) that minimizes the cross-validated variance estimate.

Our pre-specified, candidate adjustment variables consist of sex, age, country, being new to care at enrollment, baseline mobility (number of nights away from the community in the past 12 months), baseline viral suppression status, and nothing (i.e., unadjusted). Our pre-specified, candidate learners consist of generalized linear models (GLMs) adjusting for a single variable beyond the intervention indicator, stepwise regression, multivariate adaptive regression splines (MARS), MARS after screening based on outcome correlations, and the arm-specific mean outcome.

Primary effect estimates will be for the study sample and on the **ratio scale**: $1/n \sum_{i=1}^{n} Y_i(1) \div 1/n \sum_{i=1}^{n} Y_i(0)$, where $Y_i(1)$ denotes the counterfactual outcome for participant *i* under the intervention and $Y_i(0)$ denotes the counterfactual outcome for participant *i* under the control.[9–11] Secondary comparisons will be on the difference scale.

We will test the **null hypothesis** that the intervention did not improve viral suppression at week-48 using a one-sided test at the 5% significance level. We will also report point estimates and 95% confidence intervals for each effect measure and the arm-specific outcomes at baseline and week-48. Standard error estimation will be based on the estimated influence curve, and statistical inference will follow from the Central Limit Theorem (i.e., using the standard normal distribution).[1]

Secondary analyses: To assess the robustness of these findings, we will repeat these analyses using the unadjusted effect estimator. We will also repeat these analyses excluding persons with missing endpoints and using TMLE to adjust for missing endpoints.

Subgroup analyses: We will repeat these analyses within strata defined by country, sex, age group, mobility level, alcohol use, baseline viral suppression status, and trial enrollment criteria. In subgroups analyses, we will limit the candidate estimation approaches to main terms adjustment for a single covariate or the simple mean and use leave-one-out cross-validation for subgroups with <40 participants. To further understand effect heterogeneity, we may conduct variable importance measures (unadjusted and adjusted) to understand baseline predictors of endline viremia, overall and by arm.

Secondary endpoints compared by arm: We will implement analogous analyses to evaluate the intervention effect on retention in care and ART possession.

5. Intervention Implementation

Within the intervention arm, we will describe coverage and uptake of the intervention at baseline (week-0), week-12, week-24, and week-36 study visits:

- Visit coverage: number and proportion who attended study visits
- Choice of intervention components: number and proportion* who selected the travel pack, longer ART refills, off-site refills, and/or facilitated out-transfer
- Stocking and re-stocking of travel pack components: number and proportion* who selected the hotline, emergency ART, the travel checklist, and/or discrete ART packaging
- Choice of visit type: number and proportion who had visits in-person at the clinic, in-person in the community, or virtually on the mobile hotline

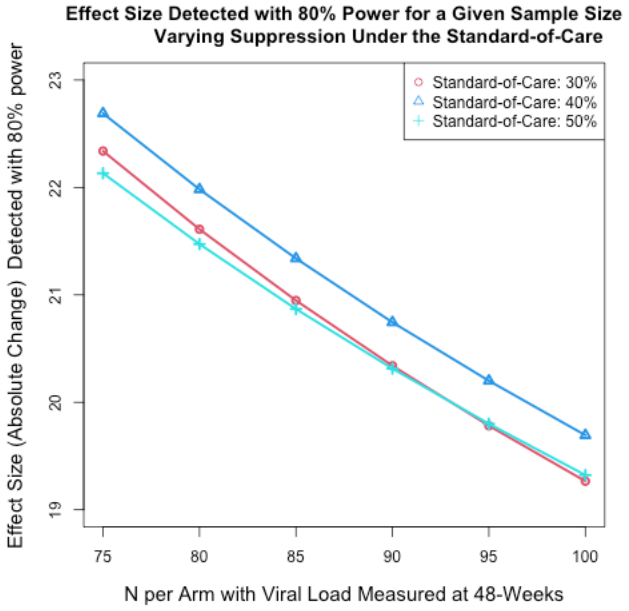
*Since participants could select more than one component, the proportion for a given visit-week may exceed 100%. Additionally, for unscheduled visits, we will report analogous metrics in terms of ever choice or use. We may also report summaries of the travel screen conducted by the mobility coordinator at each visit.

Appendix: Power calculations

Sample size and power calculations were based on a two-sample test of proportions with *power.prop.test* function in *R*.[12] We expect these calculations to be conservative,

because of the precision gained through stratified randomization, covariate adjustment during the analysis, and our pre-specified use of a one-sided hypothesis test.

We estimated 100 participants/arm would provide 80% power to detect at least a 20% absolute increase in viral suppression at 48-weeks from 40% under the standard-of-care. As shown in the following Figure, even with 25% attrition (from 100 to 75 participants/arm) and lower or higher than suppression in the control, these calculations suggest we would be well-powered to detect at least a 22.3% absolute increase in viral suppress.



References

1. van der Laan M, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer; 2011.

2. Moore KL, van der Laan MJ. Covariate Adjustment in Randomized Trials with Binary Outcomes: Targeted Maximum Likelihood Estimation. *Stat Med*. 2009;28(1):39-64. doi:10.1002/sim.3445

3. Rosenblum M, van der Laan MJ. Simple, Efficient Estimators of Treatment Effects in Randomized Trials Using Generalized Linear Models to Leverage Baseline Variables. *Int J Biostat*. 2010;6(1):Article 13. doi:10.2202/1557-4679.1138

4. van der Laan MJ, Rose S. *Targeted Learning in Data Science*. Springer; 2018.

5. Benkeser D, Díaz I, Luedtke A, Segal J, Scharfstein D, Rosenblum M. Improving precision and power in randomized trials for COVID-19 treatments using covariate adjustment, for binary, ordinal, and time-to-event outcomes. *Biometrics*. 2021;n/a(n/a):1-15. doi:10.1111/biom.13377

6. Balzer L, van der Laan MJ, Petersen M, SEARCH Collaboration. Adaptive Pre-specification in Randomized Trials With and Without Pair-Matching. *Stat Med*. 2016;35(10):4528-4545. doi:10.1002/sim.7023

7. Havlir DV, Balzer LB, Charlebois ED, et al. HIV Testing and Treatment with the Use of a Community Health Approach in Rural Africa. *N Engl J Med*. 2019;381(3):219-229.

8. Balzer LB, van der Laan M, Ayieko J, et al. Two-Stage TMLE to Reduce Bias and Improve Efficiency in Cluster Randomized Trials. *Biostatistics*. 2021;kxab043. https://doi.org/10.1093/biostatistics/kxab043

9. Neyman J. Sur les applications de la theorie des probabilites aux experiences agricoles: Essai des principes (In Polish). English translation by D.M. Dabrowska and T.P. Speed (1990). *Stat Sci*. 1923;5:465-480.

10. Rubin DB. Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Stat Sci*. 1990;5(4):472-480.

11. Balzer LB, Petersen ML, van der Laan MJ. Targeted estimation and inference of the sample average treatment effect in trials with and without pair-matching. *Stat Med*. 2016;35(21):3717-3732. doi:10.1002/sim.6965

12. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2022. http://www.R-project.org

Sustainable East Africa Research in Community Health (SEARCH) Collaboration

Statistical Analysis Plan for Dynamic Choice Care Intervention for Persons with HIV and Unhealthy Alcohol Use in Phase A of SEARCH-Sapphire

Laura B. Balzer, PhD[1]
and the SEARCH Study Team[2]

May 19, 2022

v1.0

[1]University of California, Berkeley
[2]www.searchendaids.com

Table of Contents:

1. Study Overview

In Phase A of SEARCH-Sapphire (NCT04810650), we are conducting an individually randomized controlled trial to evaluate the effect of an alcohol counseling intervention for persons with HIV (PWH) and unhealthy alcohol use in rural Kenya and Uganda. Details of the trial design and procedures can be found in the corresponding Study Protocol. Analyses plans for qualitative outcomes and cost-effectiveness outcomes are available elsewhere. Power calculations are given in the Appendix.

In brief, from April through September 2021, we 401 enrolled PWH with unhealthy alcohol use and risk of viral non-suppression. These participants were randomized to the intervention or the standard-of-care using a stratified random block design with stratification factors of country and sex and with random block sizes of 2 and 4. The randomization list was generated by an independent researcher.

The intervention is culturally-adapted and skills-based alcohol counseling, consisting of in-person sessions at baseline and every 12 weeks as well as phone-based “booster” calls every 3 weeks between in-person sessions. The intervention is delivered by lay counselors trained by a licensed clinical psychologist, who also provides ongoing supervision and feedback.

The primary objective is to evaluate if the intervention improved viral suppression (<400 copies/mL) among PWH with unhealthy alcohol use after 24 weeks of follow-up. Secondary endpoints, compared between randomized arms at 24-weeks, include alcohol use, as measured by Alcohol Use Disorders Identification Test–

Consumption [AUDIT-C] or phosphatidylethanol [PEth]. Additionally, within the intervention arm, we will report completion of counseling sessions and their fidelity.

2. Population and Characteristics

The population of interest is PWH who are

- Aged 18+ years
- Enrolled or new to HIV care
- At risk of viral non-suppression:
    - HIV RNA>400 copies/mL in the past 12 months
    - Missed visit (>2 weeks and <90 days from scheduled appointment) in the past 6 months
    - Re-engaging in care (>90 days from last scheduled visit) in the past 6 months
    - New HIV diagnosis (not yet on ART or started ART within past 1 month)
- Have unhealthy alcohol use: AUDIT-C over the past 3-months of ≥3 for women or ≥4 for men

To characterize measurement of this population, we will provide a participant flow diagram (i.e., a CONSORT diagram). Overall and stratified by trial arm and further by sex, we will summarize the baseline characteristics, including sex, age, country, marital status, occupation, education level, literacy level, trial recruitment site (i.e., venue-based or clinic), trial enrollment criteria, baseline ART status and regimen, baseline viral suppression status, and baseline alcohol use via AUDIT-C score and PEth levels. We will categorize age as “younger” if aged 18-30 years. When treated continuously, PEth will be log-10 transformed and values below the limit of quantification (8 ng/mL) set to 0.

3. Endpoint Measurement and Definition

The **primary endpoint** is HIV viral suppression (HIV RNA<400 copies/mL) at 24-weeks. The primary analytic population will consist of all study participants residing in the study region. In other words, the primary analysis will exclude persons who died, withdrew, or moved out of the study region. In the primary analysis, missing endpoint viral loads will be treated as failures (i.e., unsuppressed). In pre-specified sensitivity analyses, described below, we will assess the robustness of these analytic choices.

We will also examine the following **secondary endpoints** to capture alcohol use at 24-weeks:

- Composite measure of unhealthy alcohol use: AUDIT-C ≥3 for women and ≥4 for men or PEth ≥50 ng/mL
- AUDIT-C ≥3 for women and ≥4 for men
- PEth ≥ 50ng/mL
- Composite measure of unhealthy alcohol use: AUDIT-C ≥6 or PEth >200 ng/mL
- AUDIT-C ≥6
- PEth >200 ng/mL
- Log-10 transformed PEth (treated continuously)

Pending data availability, we will define the primary and secondary endpoints analogously at 48-weeks.

4. Evaluation of the SEARCH Intervention Effect

We will assess the intervention effect with targeted minimum loss-based estimation (TMLE), which improves precision and power by adaptively adjusting for baseline outcome predictors.[1–5] Here, we will use **TMLE with Adaptive Pre-specification** to flexibly control for baseline covariates, while maintaining Type-I error control and accounting for the randomization scheme.[6–8] Using 10-fold cross-validation, we will chose the optimal approach for estimating the outcome regression (i.e., the expected outcome given the randomization arm and adjustment covariates) and the known propensity score (i.e., the conditional probability of being randomized to the intervention given the adjustment covariates). Specifically, we will select the combination of estimators (adjustment variables + approach) that minimizes the cross-validated variance estimate.

Our pre-specified, candidate adjustment variables consist of sex, age, country, baseline viral suppression status, baseline PEth (log-10 transformed), baseline AUDIT-C score, and nothing (i.e., unadjusted). Our pre-specified, candidate learners consist of generalized linear models (GLMs) adjusting for a single variable beyond the intervention indicator, stepwise regression, multivariate adaptive regression splines (MARS), MARS after screening based on outcome correlations, and the arm-specific mean outcome.

Primary effect estimates will be for the study sample and on the **ratio scale**: $1/n \sum_{i=1}^{n} Y_i(1) \div 1/n \sum_{i=1}^{n} Y_i(0)$, where $Y_i(1)$ denotes the counterfactual outcome for participant *i* under the intervention and $Y_i(0)$ denotes the counterfactual outcome for participant *i* under the control.[9–11] Secondary comparisons will be on the difference scale.

We will test the **null hypothesis** that the intervention did not improve viral suppression at 24-weeks using a one-sided test at the 5% significance level. We will also report point estimates and 95% confidence intervals for each effect measure and the arm-specific outcomes at baseline and 24-weeks. Standard error estimation will be based on the estimated influence curve, and statistical inference will follow from the Central Limit Theorem (i.e., using the standard normal distribution).[1]

Secondary analyses: To assess the robustness of these findings, we will repeat these analyses using the unadjusted effect estimator. We will also repeat these analyses including persons who moved out of the study region, excluding persons with missing endpoints, and using TMLE to adjust for missing endpoints.

Subgroup analyses: We will repeat these analyses within strata defined by country, sex, age group, trial recruitment site, trial enrollment criteria, baseline viral suppression status, and baseline alcohol use: AUDIT-C score 3-5/women or 4-5/men, AUDIT-C score 6-7, AUDIT-C score 8+, PEth<8 ng/mL (undetectable), PEth≥8 ng/mL, PEth<50 ng/mL, PEth≥50 ng/mL, PEth≤200 ng/mL, Peth>200 ng/mL. In subgroups analyses, we will limit the candidate estimation approaches to main terms adjustment for a single covariate or the simple mean and use leave-one-out cross-validation for subgroups with <40 participants. To further understand effect heterogeneity, we may conduct variable importance measures (unadjusted and adjusted) to understand baseline predictors of endline viremia, overall and by arm.

Secondary endpoints compared by arm: We will implement analogous analyses to evaluate the intervention effect on alcohol use. In these analyses, the primary approach will exclude persons with missing measures of alcohol use at 24-weeks. We will again examine the robustness of this approach (i.e., adjust with TMLE for missing measures and include persons who moved out of the study region). When examining the effect on log-10 PEth values, we will use the difference scale.

5. Intervention Fidelity and Implementation

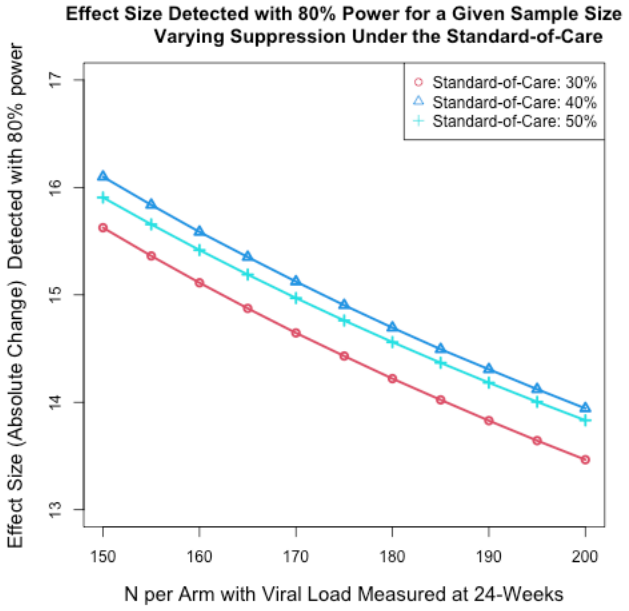
Within the intervention arm, we will evaluations of content fidelity and counseling skills at randomly selected in-person sessions. We will also report intervention adherence, as the proportion of participants who completed each counseling session.

Appendix: Power calculations

Sample size and power calculations were based on a two-sample test of proportions with *power.prop.test* function in *R*.[12] We expect these calculations to be conservative,

because of the precision gained through stratified randomization, covariate adjustment during the analysis, and our pre-specified use of a one-sided hypothesis test.

We estimated 200 participants/arm would provide 80% power to detect at least a 14% absolute increase in viral suppression at 24-weeks from 40% under the standard-of-care. As shown in the following Figure, even with 25% attrition (from 200 to 150 participants/arm) and lower or higher than suppression in the control, these calculations suggest we would be well-powered to detect at least a 16% absolute increase in viral suppress.



References

1. van der Laan M, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer; 2011.

2. Moore KL, van der Laan MJ. Covariate Adjustment in Randomized Trials with Binary Outcomes: Targeted Maximum Likelihood Estimation. *Stat Med*. 2009;28(1):39-64. doi:10.1002/sim.3445

3. Rosenblum M, van der Laan MJ. Simple, Efficient Estimators of Treatment Effects in Randomized Trials Using Generalized Linear Models to Leverage Baseline Variables. *Int J Biostat*. 2010;6(1):Article 13. doi:10.2202/1557-4679.1138

4. van der Laan MJ, Rose S. *Targeted Learning in Data Science*. Springer; 2018.

5. Benkeser D, Díaz I, Luedtke A, Segal J, Scharfstein D, Rosenblum M. Improving precision and power in randomized trials for COVID-19 treatments using covariate adjustment, for binary, ordinal, and time-to-event outcomes. *Biometrics*. 2021;n/a(n/a):1-15. doi:10.1111/biom.13377

6. Balzer L, van der Laan MJ, Petersen M, SEARCH Collaboration. Adaptive Pre-specification in Randomized Trials With and Without Pair-Matching. *Stat Med*. 2016;35(10):4528-4545. doi:10.1002/sim.7023

7. Havlir DV, Balzer LB, Charlebois ED, et al. HIV Testing and Treatment with the Use of a Community Health Approach in Rural Africa. *N Engl J Med*. 2019;381(3):219-229.

8. Balzer LB, van der Laan M, Ayieko J, et al. Two-Stage TMLE to Reduce Bias and Improve Efficiency in Cluster Randomized Trials. *Biostatistics*. 2021;kxab043. https://doi.org/10.1093/biostatistics/kxab043

9. Neyman J. Sur les applications de la theorie des probabilites aux experiences agricoles: Essai des principes (In Polish). English translation by D.M. Dabrowska and T.P. Speed (1990). *Stat Sci*. 1923;5:465-480.

10. Rubin DB. Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Stat Sci*. 1990;5(4):472-480.

11. Balzer LB, Petersen ML, van der Laan MJ. Targeted estimation and inference of the sample average treatment effect in trials with and without pair-matching. *Stat Med*. 2016;35(21):3717-3732. doi:10.1002/sim.6965

12. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2022. http://www.R-project.org

Statistical Analysis Plan for Dynamic HIV Choice Prevention delivered with Community Health Workers in Phase A of SEARCH-Sapphire

Laura B. Balzer[1], Elijah R. Kakande[2], James Ayieko[3], Jane Kabami[2,4], Gabriel Chamie[5], Moses R. Kamya[2,4], Diane V. Havlir[5], Maya L. Petersen[1] for the SEARCH Collaboration[6]

Abstract:
This document provides the statistical analytic plan (SAP) for a cluster randomized trial evaluating the effect of community-based Dynamic Choice HIV Prevention (DCP) intervention, delivered by community health workers in rural Uganda and Kenya (Clinicaltrials.gov: NCT04810650). The SAP was locked prior to unblinding and effect estimation.

July 11, 2022

v1.0

1 Division of Biostatistics, University of California Berkeley, Berkeley, USA
2 Infectious Diseases Research Collaboration, Kampala, Uganda
3 Kenya Medical Research Institute, Nairobi, Kenya
4 Department of Medicine, Makerere University College of Health Sciences, Kampala, Uganda
5 Division of HIV, Infectious Diseases, and Global Medicine, University of California San Francisco, San Francisco, USA
[6]www.searchendaids.com

Table of Contents:

1. Study Overview

In Phase A of SEARCH-Sapphire (NCT04810650), we are conducting a cluster randomized controlled trial to evaluate the effect of Dynamic Choice HIV Prevention (DCP) delivered by community health workers on biomedical HIV prevention coverage in rural Kenya and Uganda. Details of the trial design and procedures can be found in the corresponding Study Protocol. Analysis plans for qualitative outcomes and cost-effectiveness outcomes are available elsewhere. Power calculations are given in the Appendix.

In brief, we selected 16 villages in rural settings with substantial HIV risk in Kenya and Uganda. These were pair-matched within country on size of village, number of community health workers, and proximity to the highway or a trading center. Then villages were randomized within matched pairs to the DCP intervention or the standard-of-care. Randomization took place at community-based events, where community leaders selected from sealed envelopes containing the trial arm. From May-August 2021, we screened and enrolled 429 participants who were currently or anticipated being at risk of HIV. Follow-up is over 48 weeks.

The DCP intervention includes choice of HIV prevention product (oral pre-exposure prophylaxis [PrEP] or post-exposure prophylaxis [PEP]), choice in HIV testing, choice in service location, and provider training on patient-centered care.

The primary objective is to evaluate if the DCP intervention improved biomedical HIV prevention coverage, defined as the proportion of the follow-up where the participant self-reported using PrEP or PEP. Secondary endpoints include

biomedical covered time during periods of self-assessed HIV risk (compared between randomized arms) as well as coverage and uptake of the DCP intervention components (within intervention arm only).

2. Population and Characteristics

The population of interest is persons aged 15+ years who are residing in study villages, HIV-negative by country-standard rapid testing algorithm, and reporting HIV risk, as assessed via the country-specific Ministry of Health screening tools or self-assessed.

To characterize measurement of this population, we will provide a participant flow diagram (i.e., a CONSORT diagram). Overall and stratified by trial arm and further by sex, we will summarize the baseline characteristics, including sex, age, country, marital status, occupation, HIV risk criteria, alcohol use (any in prior 3 months), mobility (nights away in the past 3 months), pregnancy (women only), circumcision (men only), and any prior use of PrEP or PEP in the past 6 months. We will discretize age into “younger” if aged 15-24 years or “older” if aged 25+ years.

3. Endpoint Measurement and Definition

At week-24 and week-48 of follow-up, surveys will be administered to assess HIV risk, possession of PrEP pills, possession of PEP pills, use of PrEP (any doses taken), and use of PEP (any doses taken). The assessment is by month and covers the prior 6 months. We will visualize these data with heatmaps and describe changes in product use over time and by self-reported risk.

The **primary endpoint of biomedical HIV prevention coverage** (a.k.a., biomedical covered time) is the proportion of follow-up where the participant reports taking PrEP or PEP. Thereby, this endpoint has a minimum of 0% (no use) and a maximum of 100% (full coverage). Persons contribute follow-up time when they respond to a survey. Persons who fail to complete both week-24 and week-48 surveys are missing in the primary analysis. Persons with incident HIV infection are assumed not to be covered during the period prior to seroconversion.

Using these data, we will also define the following **secondary endpoints**:

- Biomedical covered time at-risk, where follow-up is restricted to months of self-reported risk
- Possession covered time, defined as the proportion of follow-up where the participant reports having or receiving PrEP or PEP pills

- Possession covered time at-risk
- Use-to-possession ratio, defined as the proportion of follow-up with PrEP or PEP pills where the participant reports taking them
- Use-to-possession ratio when at-risk

4. Evaluation of the SEARCH DCP Intervention Effect

We will assess the Dynamic Choice Prevention intervention effect with targeted minimum loss-based estimation (TMLE), which improves precision and power by adaptively adjusting for baseline outcome predictors.[1–5] Here, we will use **TMLE with Adaptive Pre-specification** to flexibly control for baseline covariates, while maintaining Type-I error control and accounting for clustering.[6–8] Using leave-one-village-out cross-validation, we will chose the optimal approach for estimating the outcome regression (i.e., the expected outcome given the randomization arm and adjustment covariates) and the known propensity score (i.e., the conditional probability of being randomized to the intervention given the adjustment covariates). Specifically, we will select the combination of estimators (adjustment variables + approach) that minimizes the cross-validated variance estimate, which again accounts for clustering.

Our pre-specified, candidate adjustment variables consist of sex, age, country, use of PrEP/PEP in the 6 months prior to enrollment, and nothing (i.e., unadjusted). Our pre-specified, candidate learners consist of generalized linear models (GLMs) adjusting for a single variable (beyond the intervention indicator), stepwise regression, multivariate adaptive regression splines (MARS), and the arm-specific mean outcome.

Primary effect estimates will be for the study sample and on the **difference scale**: $1/n \sum_{i=1}^{n}[Y_i(1) - Y_i(0)]$, where n denotes the total participants, $Y_i(1)$ denotes the counterfactual outcome for participant *i* under the intervention and $Y_i(0)$ denotes the counterfactual outcome for participant *i* under the control.[9–11] In other words, we will estimate an individual-level effect (i.e., weight the n individuals equally); secondary analyses will estimate a cluster-level effect (i.e., weight the J villages equally regardless of their size).[12–14] Secondary comparisons will be on the ratio scale. All analyses will break the matches used for randomization.[15]

We will test the **null hypothesis** that the intervention did not change biomedical covered time with a two-sided test at the 5% significance level. We will also report point estimates and 95% confidence intervals for each effect measure and the arm-specific outcomes. Standard error estimation will be based on the estimated influence curve, appropriately aggregated to the cluster-level.[14,16,17] Statistical inference will follow from

the Central Limit Theorem,[1] and will use the Student's *t*-distribution with $(J-2)$ as a finite sample approximation to the standard normal distribution.[15]

Secondary analyses: To assess the robustness of these findings, we will repeat these analyses using the unadjusted effect estimator. We will also repeat these analyses using TMLE to adjust for missing endpoints.

Subgroup analyses: We will repeat these analyses within strata defined by country, sex, age group, alcohol use, and use of PrEP/PEP in the 6 months before enrollment. To further understand effect heterogeneity, we may conduct variable importance measures (i.e., unadjusted and adjusted predictor analyses) with TMLE.

Secondary endpoints compared by arm: We will implement analogous analyses to evaluate the intervention effect on biomedical covered time at-risk, possession covered time (overall and at-risk), and use-to-possession ratio (overall and at-risk).

Validation of self-report: Since the primary and key secondary study endpoints rely on self-report, we will objectively measure adherence using drug levels in small hair samples collected among participants reporting any PrEP or PEP doses taken in the past 30 days. Overall and by arm, we will report the number and proportion of these participants with detectable tenofovir levels (>0.002 ng/mg) in their hair at week-24. Using a two-sample test, we will formally test the null hypothesis of equal proportions between arms. We may repeat these analyses at week-48.

Additional descriptive analyses: Overall and by arm, we will report the number and proportion of participants who withdrew, died, or seroconverted. We will provide seroconversion narratives and may test the null hypothesis of the HIV incidence rate is the same between arms through Poisson regression with person-years-at-risk as offset.

5. Intervention Implementation

Within the intervention arm, we will describe coverage and uptake of the DCP intervention over follow-up:

- Visit coverage: number and proportion who attended study visits
- Choice of HIV prevention product: number and proportion who selected PrEP, PEP, condoms, or nothing
- Choice of HIV testing: number and proportion who selected a self-test or rapid HIV test
- Choice of service location: number and proportion who selected to have visits at clinic or at an out-of-clinic location (e.g., home)

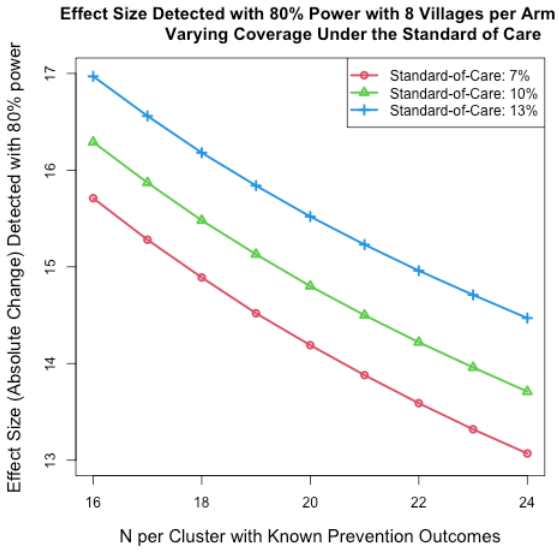
All metrics will exclude persons who died, withdrew, or seroconverted by that week of follow-up. At week-48, we will additionally exclude persons who did not reconsent to the extension study (see NCT05549726 for details).

We will also characterize ever use of DCP intervention components over follow-up. We may also report on reasons for product changes, barriers to care, plans to address those barriers, and utilization of the phone hotline. We will report these metrics overall and within key subgroups, such as sex.

Appendix: Power calculations

Sample size and power calculations were based on standard formulas for cluster randomized trials with a continuous outcome.[15] We expect these calculations to be conservative, because of the precision gained through covariate adjustment during the analysis.

We anticipate an average (harmonic mean) of 20 participants per village. Assuming 10% coverage under the standard-of-care, a standard deviation of 0.35, and coefficient of variation of k_m=0.25, we anticipate 80% power to detect at least a 15% absolute increase in prevention coverage with J=16 clusters (J=8 clusters/arm). Even with 20% fewer participants (from 20 to 16 participants/cluster), these calculations suggest we would be well-powered to detect at least a 16.5% absolute increase in prevention coverage.



References

1. van der Laan M, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer; 2011.

2. Moore KL, van der Laan MJ. Covariate Adjustment in Randomized Trials with Binary Outcomes: Targeted Maximum Likelihood Estimation. *Stat Med*. 2009;28(1):39-64. doi:10.1002/sim.3445

3. Rosenblum M, van der Laan MJ. Simple, Efficient Estimators of Treatment Effects in Randomized Trials Using Generalized Linear Models to Leverage Baseline Variables. *Int J Biostat*. 2010;6(1):Article 13. doi:10.2202/1557-4679.1138

4. van der Laan MJ, Rose S. *Targeted Learning in Data Science*. Springer; 2018.

5. Benkeser D, Díaz I, Luedtke A, Segal J, Scharfstein D, Rosenblum M. Improving precision and power in randomized trials for COVID-19 treatments using covariate adjustment, for binary, ordinal, and time-to-event outcomes. *Biometrics*. 2021;n/a(n/a):1-15. doi:10.1111/biom.13377

6. Balzer L, van der Laan MJ, Petersen M, SEARCH Collaboration. Adaptive Pre-specification in Randomized Trials With and Without Pair-Matching. *Stat Med*. 2016;35(10):4528-4545. doi:10.1002/sim.7023

7. Havlir DV, Balzer LB, Charlebois ED, et al. HIV Testing and Treatment with the Use of a Community Health Approach in Rural Africa. *N Engl J Med*. 2019;381(3):219-229.

8. Balzer LB, van der Laan M, Ayieko J, et al. Two-Stage TMLE to Reduce Bias and Improve Efficiency in Cluster Randomized Trials. *Biostatistics*. 2021;kxab043. https://doi.org/10.1093/biostatistics/kxab043

9. Neyman J. Sur les applications de la theorie des probabilites aux experiences agricoles: Essai des principes (In Polish). English translation by D.M. Dabrowska and T.P. Speed (1990). *Stat Sci*. 1923;5:465-480.

10. Rubin DB. Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Stat Sci*. 1990;5(4):472-480.

11. Balzer LB, Petersen ML, van der Laan MJ. Targeted estimation and inference of the sample average treatment effect in trials with and without pair-matching. *Stat Med*. 2016;35(21):3717-3732. doi:10.1002/sim.6965

12. Balzer LB, Zheng W, van der Laan MJ, Petersen ML. A new approach to hierarchical data analysis: Targeted maximum likelihood estimation for the causal effect of a cluster-level exposure. *Stat Methods Med Res*. 2019;28(6):1761-1780. doi:10.1177/0962280218774936

13.Imai K, King G, Nall C. The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation. *Stat Sci*. 2009;24(1):29-53. doi:10.1214/08-STS274

14.Benitez A, Petersen ML, van der Laan MJ, et al. Defining and Estimating Effects in Cluster Randomized Trials: A Methods Comparison. Published online January 19, 2022. doi:10.48550/arXiv.2110.09633

15.Hayes RJ, Moulton LH. *Cluster Randomised Trials*. Chapman & Hall/CRC; 2009.

16.Petersen ML, Schwab J, Gruber S, Blaser N, Schomaker M, Laan MJ van der. Targeted Maximum Likelihood Estimation for Dynamic and Static Longitudinal Marginal Structural Working Models. *J Causal Inference*. 2014;2(2). doi:10.1515/jci-2013-0007

17.Schnitzer ME, Laan MJ van der, Moodie EE, Platt RW. Effect of breastfeeding on gastrointestinal infection in infants: a targeted maximum likelihood approach for clustered longitudinal data. *Ann Appl Stat*. 2014;8(2):703-725.

Sustainable East Africa Research in Community Health (SEARCH) Collaboration

Statistical Analysis Plan for Dynamic HIV Choice Prevention at Outpatient Departments in Phase A of SEARCH-Sapphire

Laura B. Balzer, PhD[1]
and the SEARCH Study Team[2]

July 11, 2022

v1.0

[1]University of California, Berkeley
[2]www.searchendaids.com

Table of Contents:

1. Study Overview

In Phase A of SEARCH-Sapphire (NCT04810650), we are conducting an individually randomized controlled trial to evaluate the effect of Dynamic Choice HIV Prevention (DCP) delivered at Outpatient Departments on biomedical HIV prevention coverage in rural Kenya and Uganda. Details of the trial design and procedures can be found in the corresponding Study Protocol. Analyses plans for qualitative outcomes and cost-effectiveness outcomes are available elsewhere. Power calculations are given in the Appendix.

In brief, from April-July 2021, we enrolled 403 participants who were currently or anticipated being at risk of HIV. These participants were randomized to the DCP intervention or the standard-of-care using a stratified random block design with stratification factors of country and sex and with random block sizes of 2 and 4. The randomization list was generated by an independent researcher.

The DCP intervention includes choice of HIV prevention product (oral pre-exposure prophylaxis [PrEP] or post-exposure prophylaxis [PEP]), choice in HIV testing, choice in service location, and provider training on patient-centered care. Follow-up is over 48 weeks.

The primary objective is to evaluate if the DCP intervention improved biomedical HIV prevention coverage, defined as the proportion of the follow-up where the participant self-reported using PrEP or PEP. Secondary endpoints include biomedical covered time during periods of self-assessed HIV risk (compared between randomized arms) as well as coverage and uptake of the DCP intervention components (within intervention arm only).

2. Population and Characteristics

The population of interest is persons aged 15+ years who are seen at Outpatient Departments and report current or anticipated HIV risk, as assessed via the country-specific Ministry of Health screening tools or self-assessed.

To characterize measurement of this population, we will provide a participant flow diagram (i.e., a CONSORT diagram). Overall and stratified by trial arm and further by sex, we will summarize the baseline characteristics, including sex, age, country, marital status, occupation, HIV risk criteria, alcohol use (any in prior 3 months), mobility (nights away in the past 3 months), pregnancy (women only), circumcision (men only), and any prior use of PrEP or PEP in the past 6 months. We will categorize age in “younger” if aged 15-24 years.

3. Endpoint Measurement and Definition

At week-24 and week-48 of follow-up, surveys will be administered to assess HIV risk, possession of PrEP pills, possession of PEP pills, use of PrEP (any doses taken), and use of PEP (any doses taken). The assessment is by month and covers the prior 6 months. We will visualize these data with heatmaps and describe changes in product use over time and by self-reported risk.

The **primary endpoint of biomedical HIV prevention coverage** (a.k.a., biomedical covered time) is the proportion of follow-up where the participant reports taking PrEP or PEP. Thereby, this endpoint has a minimum of 0% (no use) and a maximum of 100% (full coverage). Persons contribute follow-up time when they respond to a survey. Persons who fail to complete both week-24 and week-48 surveys are missing in the primary analysis. Persons with incident HIV infection are assumed not to be covered during the period prior to seroconversion.

Using these data, we will also define the following **secondary endpoints**:

- Biomedical covered time at-risk, where follow-up is restricted to months of self-reported risk
- Possession covered time, defined as the proportion of follow-up where the participant reports having or receiving PrEP or PEP pills
- Possession covered time at-risk
- Use/possession ratio, defined as the proportion of follow-up with PrEP or PEP pills where the participant reports taking them
- Use/possession ratio when at-risk

4. Evaluation of the SEARCH DCP Intervention Effect

We will assess the Dynamic Choice Prevention intervention effect with targeted minimum loss-based estimation (TMLE), which improves precision and power by adaptively adjusting for baseline outcome predictors.[1–5] Here, we will use **TMLE with Adaptive Pre-specification** to flexibly control for baseline covariates, while maintaining Type-I error control and accounting for the randomization scheme.[6–8] Using 10-fold cross-validation, we will chose the optimal approach for estimating the outcome regression (i.e., the expected outcome given the randomization arm and adjustment covariates) and the known propensity score (i.e., the conditional probability of being randomized to the intervention given the adjustment covariates). Specifically, we will select the combination of estimators (adjustment variables + approach) that minimizes the cross-validated variance estimate.

Our pre-specified, candidate adjustment variables consist of sex, age, country, use of PrEP/PEP before enrollment, and nothing (i.e., unadjusted). Our pre-specified, candidate learners consist of generalized linear models (GLMs) adjusting for a single variable (beyond the intervention indicator), stepwise regression, multivariate adaptive regression splines (MARS), and the arm-specific mean outcome.

Primary effect estimates will be for the study sample and on the **difference scale**: $1/n \sum_{i=1}^{n}[Y_i(1) - Y_i(0)]$, where $Y_i(1)$ denotes the counterfactual outcome for participant *i* under the intervention and $Y_i(0)$ denotes the counterfactual outcome for participant *i* under the control.[9–11] Secondary comparisons will be on the ratio scale.

We will test the **null hypothesis** that the intervention did not change biomedical covered time with a two-sided test at the 5% significance level. We will also report point estimates and 95% confidence intervals for each effect measure and the arm-specific outcomes. Standard error estimation will be based on the estimated influence curve, and statistical inference will follow from the Central Limit Theorem (i.e., using the standard normal distribution).[1]

Secondary analyses: To assess the robustness of these findings, we will repeat these analyses using the unadjusted effect estimator. We will also repeat these analyses using TMLE to adjust for missing endpoints.

Subgroup analyses: We will repeat these analyses within strata defined by country, sex, age group, alcohol use, and use of PrEP/PEP in the 6 months before enrollment. In subgroups with fewer than 41 participants, we will limit the candidate estimation approaches to main terms adjustment for a single covariate or the simple mean and use

leave-one-out cross-validation. To further understand effect heterogeneity, we may conduct variable importance measures (i.e., unadjusted and adjusted predictor analyses) with TMLE.

Secondary endpoints compared by arm: We will implement analogous analyses to evaluate the intervention effect on biomedical covered time at-risk, possession covered time (overall and at-risk), and use/possession ratio (overall and at-risk).

Validation of self-report: Since the primary and key secondary study endpoints rely on self-report, we will objectively measure adherence using drug levels in small hair samples collected among participants reporting any PrEP or PEP doses taken in the past 30 days. Overall and by arm, we will report the number and proportion of these participants with detectable tenofovir levels (>0.002 ng/mg) in their hair at week-24. Using a two-sample test, we will formally test the null hypothesis of equal proportions between arms. We may repeat these analyses at week-48.

Additional descriptive analyses: Overall and by arm, we will report the number and proportion of participants who withdrew, died, or seroconverted. We will provide seroconversion narratives and may test the null hypothesis of the HIV incidence rate is the same between arms through Poisson regression with person-years-at-risk as offset.

5. Intervention Implementation

Within the intervention arm, we will describe coverage and uptake of the DCP intervention at baseline (week-0), week-4, week-12, week-24, week-36, and week-48:

- Visit coverage: number and proportion who attended study visits
- Choice of prevention product: number and proportion who selected PrEP, PEP, condoms, or nothing
- Choice of HIV testing: number and proportion who selected a self-test or rapid HIV test
- Choice of service location: number and proportion who selected to have visits at clinic or at an out-of-facility location (e.g., home)

All metrics will exclude persons who died, withdrew, or seroconverted by that week of follow-up. At week-48, we will additionally exclude persons who did not reconsent to the extension study (see NCT05549726 for details).

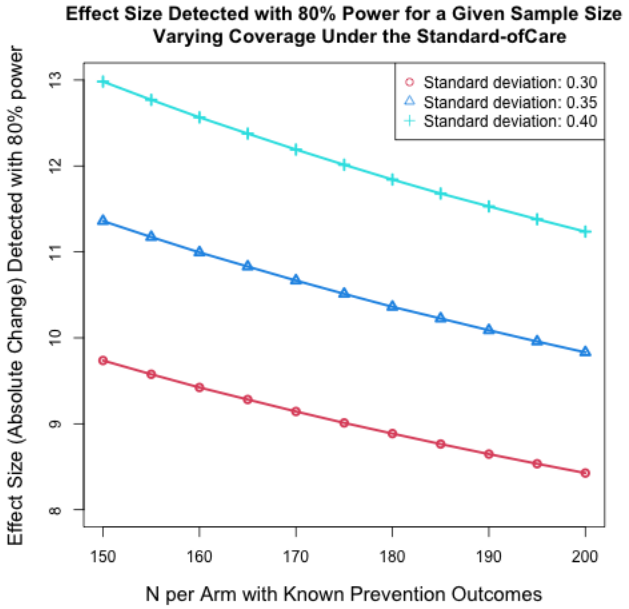
We will also characterize ever use of DCP intervention components over follow-up. We may also report on reasons for product changes, barriers to care, plans to address

those barriers, and utilization of the phone hotline. We will report these metrics overall and within key subgroups, such as sex.

Appendix: Power calculations

Sample size and power calculations were based on a two-sample t-test with *power.t.test* function in *R*.[12] We expect these calculations to be conservative, because of the precision gained through stratified randomization and covariate adjustment during the analysis.

We estimated 200 participants/arm would provide 80% power to detect at least a 10% absolute increase in biomedical HIV prevention coverage, assuming a standard deviation of 0.3. As shown in the following Figure, even with 25% attrition (from 200 to 150 participants/arm) and higher than expected variability (e.g., standard deviation=0.40), these calculations suggest we would be well-powered to detect at least a 13% absolute increase in prevention coverage.



References

1. van der Laan M, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer; 2011.

2. Moore KL, van der Laan MJ. Covariate Adjustment in Randomized Trials with Binary Outcomes: Targeted Maximum Likelihood Estimation. *Stat Med*. 2009;28(1):39-64. doi:10.1002/sim.3445

3. Rosenblum M, van der Laan MJ. Simple, Efficient Estimators of Treatment Effects in Randomized Trials Using Generalized Linear Models to Leverage Baseline Variables. *Int J Biostat*. 2010;6(1):Article 13. doi:10.2202/1557-4679.1138

4. van der Laan MJ, Rose S. *Targeted Learning in Data Science*. Springer; 2018.

5. Benkeser D, Díaz I, Luedtke A, Segal J, Scharfstein D, Rosenblum M. Improving precision and power in randomized trials for COVID-19 treatments using covariate adjustment, for binary, ordinal, and time-to-event outcomes. *Biometrics*. 2021;n/a(n/a):1-15. doi:10.1111/biom.13377

6. Balzer L, van der Laan MJ, Petersen M, SEARCH Collaboration. Adaptive Pre-specification in Randomized Trials With and Without Pair-Matching. *Stat Med*. 2016;35(10):4528-4545. doi:10.1002/sim.7023

7. Havlir DV, Balzer LB, Charlebois ED, et al. HIV Testing and Treatment with the Use of a Community Health Approach in Rural Africa. *N Engl J Med*. 2019;381(3):219-229.

8. Balzer LB, van der Laan M, Ayieko J, et al. Two-Stage TMLE to Reduce Bias and Improve Efficiency in Cluster Randomized Trials. *Biostatistics*. 2021;kxab043. https://doi.org/10.1093/biostatistics/kxab043

9. Neyman J. Sur les applications de la theorie des probabilites aux experiences agricoles: Essai des principes (In Polish). English translation by D.M. Dabrowska and T.P. Speed (1990). *Stat Sci*. 1923;5:465-480.

10. Rubin DB. Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Stat Sci*. 1990;5(4):472-480.

11. Balzer LB, Petersen ML, van der Laan MJ. Targeted estimation and inference of the sample average treatment effect in trials with and without pair-matching. *Stat Med*. 2016;35(21):3717-3732. doi:10.1002/sim.6965

12. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2022. http://www.R-project.org

Sustainable East Africa Research in Community Health (SEARCH) Collaboration

Statistical Analysis Plan for Dynamic HIV Choice Prevention at Antenatal Clinics in Phase A of SEARCH-Sapphire

Laura B. Balzer, PhD[1]
and the SEARCH Study Team[2]

July 11, 2022

v1.0

[1]University of California, Berkeley
[2]www.searchendaids.com

Table of Contents:

1. Study Overview

In Phase A of SEARCH-Sapphire (NCT04810650), we are conducting an individually randomized controlled trial to evaluate the effect of Dynamic Choice HIV Prevention (DCP) delivered to women recruited from antenatal clinics (ANC) on biomedical HIV prevention coverage in rural Kenya and Uganda. Details of the trial design and procedures can be found in the corresponding Study Protocol. Analyses plans for qualitative outcomes and cost-effectiveness outcomes are available elsewhere. Power calculations are given in the Appendix.

In brief, from April-July 2021, we enrolled 400 participants who were currently or anticipated being at risk of HIV. These participants were randomized to the DCP intervention or the standard-of-care using a stratified random block design. The stratification factors were country and pregnancy status, and the random block sizes were 2 and 4. The randomization list was generated by an independent researcher.

The DCP intervention includes choice of HIV prevention product (oral pre-exposure prophylaxis [PrEP] or post-exposure prophylaxis [PEP]), choice in HIV testing, choice in service location, and provider training on patient-centered care. Follow-up is over 48 weeks.

The primary objective is to evaluate if the DCP intervention improved biomedical HIV prevention coverage, defined as the proportion of the follow-up where the participant self-reported using PrEP or PEP. Secondary endpoints include biomedical covered time during periods of self-assessed HIV risk (compared between randomized arms) as well as coverage and uptake of the DCP intervention components (within intervention arm only).

2. Population and Characteristics

The population of interest is persons aged 15+ years who are seen at ANC and report current or anticipated HIV risk, as assessed via the country-specific Ministry of Health screening tools or self-assessment.

To characterize measurement of this population, we will provide a participant flow diagram (i.e., a CONSORT diagram). Overall and stratified by trial arm, we will summarize the baseline characteristics, including age, country, marital status, occupation, HIV risk criteria, alcohol use (any in prior 3 months), mobility (nights away in the past 3 months), pregnancy, and any prior use of PrEP or PEP in the past 6 months. We will categorize age in “younger” if aged 15-24 years.

3. Endpoint Measurement and Definition

At week-24 and week-48 of follow-up, surveys will be administered to assess HIV risk, possession of PrEP pills, possession of PEP pills, use of PrEP (any doses taken), and use of PEP (any doses taken). The assessment is by month and covers the prior 6 months. We will visualize these data with heatmaps and describe changes in product use over time and by self-reported risk.

The **primary endpoint of biomedical HIV prevention coverage** (a.k.a., biomedical covered time) is the proportion of follow-up where the participant reports taking PrEP or PEP. Thereby, this endpoint has a minimum of 0% (no use) and a maximum of 100% (full coverage). Persons contribute follow-up time when they respond to a survey. Persons who fail to complete both week-24 and week-48 surveys are missing in the primary analysis. Persons with incident HIV infection are assumed not to be covered during the period prior to seroconversion.

Using these data, we will also define the following **secondary endpoints**:

- Biomedical covered time at-risk, where follow-up is restricted to months of self-reported risk
- Possession covered time, defined as the proportion of follow-up where the participant reports having or receiving PrEP or PEP pills
- Possession covered time at-risk
- Use/possession ratio, defined as the proportion of follow-up with PrEP or PEP pills where the participant reports taking them
- Use/possession ratio when at-risk

4. Evaluation of the SEARCH DCP Intervention Effect

We will assess the Dynamic Choice Prevention intervention effect with targeted minimum loss-based estimation (TMLE), which improves precision and power by adaptively adjusting for baseline outcome predictors.[1–5] Here, we will use **TMLE with Adaptive Pre-specification** to flexibly control for baseline covariates, while maintaining Type-I error control and accounting for the randomization scheme.[6–8] Using 10-fold cross-validation, we will chose the optimal approach for estimating the outcome regression (i.e., the expected outcome given the randomization arm and adjustment covariates) and the known propensity score (i.e., the conditional probability of being randomized to the intervention given the adjustment covariates). Specifically, we will select the combination of estimators (adjustment variables + approach) that minimizes the cross-validated variance estimate.

Our pre-specified, candidate adjustment variables consist of pregnancy status at enrollment, age, country, use of PrEP/PEP in the 6 months before enrollment, and nothing (i.e., unadjusted). Our pre-specified, candidate learners consist of generalized linear models (GLMs) adjusting for a single variable (beyond the intervention indicator), stepwise regression, multivariate adaptive regression splines (MARS), and the arm-specific mean outcome.

Primary effect estimates will be for the study sample and on the **difference scale**: $1/n \sum_{i=1}^{n}[Y_i(1) - Y_i(0)]$, where $Y_i(1)$ denotes the counterfactual outcome for participant *i* under the intervention and $Y_i(0)$ denotes the counterfactual outcome for participant *i* under the control.[9–11] Secondary comparisons will be on the ratio scale.

We will test the **null hypothesis** that the intervention did not change biomedical covered time with a two-sided test at the 5% significance level. We will also report point estimates and 95% confidence intervals for each effect measure and the arm-specific outcomes. Standard error estimation will be based on the estimated influence curve, and statistical inference will follow from the Central Limit Theorem (i.e., using the standard normal distribution).[1]

Secondary analyses: To assess the robustness of these findings, we will repeat these analyses using the unadjusted effect estimator. We will also repeat these analyses using TMLE to adjust for missing endpoints.

Subgroup analyses: We will repeat these analyses within strata defined by country, pregnancy status at enrollment, age group, alcohol use, and use of PrEP/PEP in the 6 months before enrollment. In subgroups with fewer than 41 participants, we will limit the

candidate estimation approaches to main terms adjustment for a single covariate or the simple mean and use leave-one-out cross-validation. To further understand effect heterogeneity, we may conduct variable importance measures (i.e., unadjusted and adjusted predictor analyses) with TMLE.

Secondary endpoints compared by arm: We will implement analogous analyses to evaluate the intervention effect on biomedical covered time at-risk, possession covered time (overall and at-risk), and use/possession ratio (overall and at-risk).

Validation of self-report: Since the primary and key secondary study endpoints rely on self-report, we will objectively measure adherence using drug levels in small hair samples collected among participants reporting any PrEP or PEP doses taken in the past 30 days. Overall and by arm, we will report the number and proportion of these participants with detectable tenofovir levels (>0.002 ng/mg) in their hair at week-24. Using a two-sample test, we will formally test the null hypothesis of equal proportions between arms. We may repeat these analyses at week-48.

Additional descriptive analyses: Overall and by arm, we will report the number and proportion of participants who withdrew, died, or seroconverted. We will provide seroconversion narratives and may test the null hypothesis of the HIV incidence rate is the same between arms through Poisson regression with person-years-at-risk as offset.

5. Intervention Implementation

Within the intervention arm, we will describe coverage and uptake of the DCP intervention at baseline (week-0), week-4, week-12, week-24, and week-36:

- Visit coverage: number and proportion who attended study visits
- Choice of prevention product: number and proportion who selected PrEP, PEP, condoms, or nothing
- Choice of HIV testing: number and proportion who selected a self-test or rapid HIV test
- Choice of service location: number and proportion who selected to have visits at clinic or at an out-of-facility location (e.g., home)

All metrics will exclude persons who died, withdrew, or seroconverted by that week of follow-up.

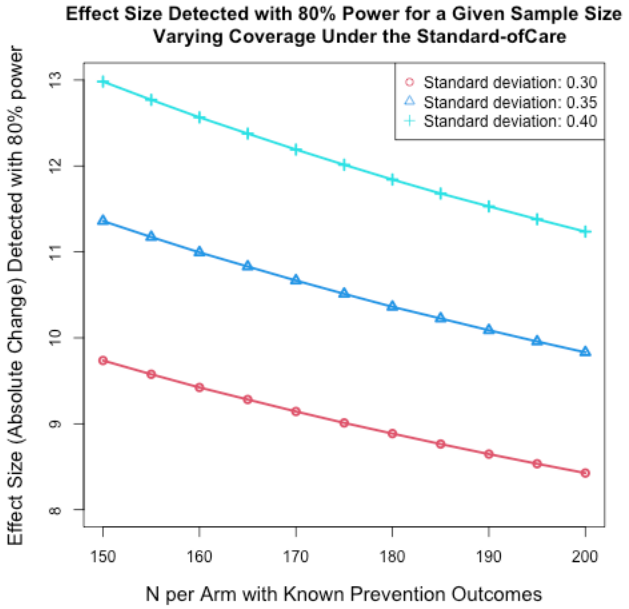
We will also characterize ever use of DCP intervention components over follow-up. We may also report on reasons for product changes, barriers to care, plans to address

those barriers, and utilization of the phone hotline. We will report these metrics overall and within key subgroups.

Appendix: Power calculations

Sample size and power calculations were based on a two-sample t-test with *power.t.test* function in *R*.[12] We expect these calculations to be conservative, because of the precision gained through stratified randomization and covariate adjustment during the analysis.

We estimated 200 participants/arm would provide 80% power to detect at least a 10% absolute increase in biomedical HIV prevention coverage, assuming a standard deviation of 0.3. As shown in the following Figure, even with 25% attrition (from 200 to 150 participants/arm) and higher than expected variability (e.g., standard deviation=0.40), these calculations suggest we would be well-powered to detect at least a 13% absolute increase in prevention coverage.



References

1. van der Laan M, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer; 2011.

2. Moore KL, van der Laan MJ. Covariate Adjustment in Randomized Trials with Binary Outcomes: Targeted Maximum Likelihood Estimation. *Stat Med*. 2009;28(1):39-64. doi:10.1002/sim.3445

3. Rosenblum M, van der Laan MJ. Simple, Efficient Estimators of Treatment Effects in Randomized Trials Using Generalized Linear Models to Leverage Baseline Variables. *Int J Biostat*. 2010;6(1):Article 13. doi:10.2202/1557-4679.1138

4. van der Laan MJ, Rose S. *Targeted Learning in Data Science*. Springer; 2018.

5. Benkeser D, Díaz I, Luedtke A, Segal J, Scharfstein D, Rosenblum M. Improving precision and power in randomized trials for COVID-19 treatments using covariate adjustment, for binary, ordinal, and time-to-event outcomes. *Biometrics*. 2021;n/a(n/a):1-15. doi:10.1111/biom.13377

6. Balzer L, van der Laan MJ, Petersen M, SEARCH Collaboration. Adaptive Pre-specification in Randomized Trials With and Without Pair-Matching. *Stat Med*. 2016;35(10):4528-4545. doi:10.1002/sim.7023

7. Havlir DV, Balzer LB, Charlebois ED, et al. HIV Testing and Treatment with the Use of a Community Health Approach in Rural Africa. *N Engl J Med*. 2019;381(3):219-229.

8. Balzer LB, van der Laan M, Ayieko J, et al. Two-Stage TMLE to Reduce Bias and Improve Efficiency in Cluster Randomized Trials. *Biostatistics*. 2021;kxab043. https://doi.org/10.1093/biostatistics/kxab043

9. Neyman J. Sur les applications de la theorie des probabilites aux experiences agricoles: Essai des principes (In Polish). English translation by D.M. Dabrowska and T.P. Speed (1990). *Stat Sci*. 1923;5:465-480.

10. Rubin DB. Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Stat Sci*. 1990;5(4):472-480.

11. Balzer LB, Petersen ML, van der Laan MJ. Targeted estimation and inference of the sample average treatment effect in trials with and without pair-matching. *Stat Med*. 2016;35(21):3717-3732. doi:10.1002/sim.6965

12. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2022. http://www.R-project.org

A Multisectoral Strategy to Address Persistent Drivers
of the HIV Epidemic in East Africa (SAPPHIRE)

Sustainable East Africa Research in Community Health (SEARCH) Consortium

Statistical Analysis Plan

Project: Effect of Community Health Worker Facilitated Telehealth Intervention for Severe Hypertension Management in SEARCH SAPPHIRE

Laura B. Balzer, PhD*
James Ayieko, MBChB, MPH, PhD**
Diane V. Havlir, MD***
Maya L. Petersen, MD, PhD*
Moses Kamya, MMed, MPH, PhD****
Matthew D. Hickey, MD***

for the SEARCH Collaboration

June 1, 2023

Version 1.0

*School of Public Health, University of California Berkeley, Berkeley, USA
**Kenya Medical Research Institute, Nairobi, Kenya
***Division of HIV, ID, & Global Medicine, Department of Medicine, UCSF, San Francisco, USA
**** Makerere University, Department of Medicine, Kampala, Uganda

Table of Contents

1. Overview

The SEARCH-Sapphire Severe Hypertension (HTN) Treatment study is an individual randomized controlled trial testing the hypothesis that community-based hypertension care (intervention) will improve control of hypertension, as compared to clinic-based, patient-centered care (control) in adults aged 40+ with severe hypertension (blood pressure ≥160/100 mmHg). The community-based intervention consists of clinician telehealth combined with lay health worker blood pressure measurement and medication delivery. The pilot trial takes place in three communities in Kenya and Uganda and began enrollment in May 2022.

After community-based screening by community health workers, individuals with severe hypertension were referred to the nearest government-run clinic for clinical assessment and enrollment. Following enrollment and completion of an initial clinic visit, participants were randomized to the intervention (n=98) or control (n=102) conditions. To ensure balance between arms, randomization was stratified by country and sex and was implemented by an independent statistician using a stratified random block design with random block sizes 2 and 4.

Intervention participants received home-based, follow-up care for hypertension, which consisted of blood pressure measurement and adherence assessment using pill count by a community health worker, a telehealth visit with a clinician, and medication dispensation by the community health worker according to clinician orders. Control participants received clinic-based integrated, patient-centered hypertension care, detailed elsewhere.[1–3] Treatment guidelines for use of antihypertensive medication for both study arms were based on standard country guidelines. Additional details of the study procedures are available in the Study Protocol.

The primary objective of this study is to determine whether the intervention improved hypertension control at 24 weeks. Secondary endpoints include retention in care, mean systolic blood pressure at 24 and 48 weeks, and hypertension control at 48 weeks. We will also characterize changes in hypertension severity and evaluate predictors HTN control and retention.

Throughout, the population of interest comprises non-pregnant adults, aged 40+ if blood pressure was elevated at both community-based and clinic measurement (≥140 mmHg systolic or ≥90 mmHg diastolic) and moderate-severely elevated on at least one of these measurements (≥160 mmHg systolic or ≥100 mmHg diastolic). Participants were excluded if they were pregnant – these individuals were immediately referred to the prenatal clinic for evaluation and treatment of hypertension in pregnancy.

We will provide descriptions of participant flow through the study (i.e., a consort diagram), measurement coverage, and baseline characteristics (e.g., country, age, sex, prior history of hypertension diagnosis and baseline treatment, comorbidities, and baseline severity of hypertension). We will also report risk factors (N and %) for uncontrolled hypertension at enrollment. We will provide these descriptive statistics, overall and by arm.

2.1 General approach for evaluating intervention effects

We will assess the intervention effect with targeted minimum loss-based estimation (TMLE), which provides precision and power gains over an unadjusted effect estimator (e.g., a t-test) by adjusting for baseline predictors of the outcome.[4–7] For a detailed review of TMLE and its relation to other effect estimators in randomized trials, we refer the reader to Colantuoni and Rosenblum.[8] For a recent demonstration of the improved precision offered by TMLE in randomized trials, we refer the reader to Balzer et al. and to Benkser et al..[9,10]

Here, we will use **TMLE with Adaptive Pre-specification**, a fully automated procedure to flexibly adjust for baseline outcome predictors, while maintaining Type-I error control.[11,12] Specifically, using 10-fold cross-validation, we will chose the optimal approach for estimating the expected outcome given the trial arm and baseline covariates (a.k.a., the outcome regression) and for estimating the conditional probability of being randomized to the intervention given the baseline covariates (a.k.a., the propensity score). Throughout, optimality is defined by using the squared (estimated) influence curve for the TMLE as loss function. Thereby, we will select the combination of estimators (adjustment variables + approach) for the outcome regression and the propensity score that minimizes the cross-validated risk estimate and, thus, maximizes empirical efficiency.

Our pre-specified, candidate adjustment variables are age, sex, baseline HTN severity as assessed at enrollment (“grade 1” as 140-159/90-99 mmHg; “grade 2” as 160-179/100-109 mmHg, and “grade 3” as ≥180/110 mmHg), country, and nothing (i.e., unadjusted). Our pre-specified candidate estimators are main terms, stepwise

regression, adjustment for a single covariate, and the mean. In sensitivity analyses, we will also implement that unadjusted effect estimator as the contrasts of average outcomes by arm.

Primary estimates will be for the study sample and on the **difference scale**: $\frac{1}{n}\sum_{i=1}^{n}[Y_i(1) - Y_i(0)]$, where $Y_i(1)$ denotes the counterfactual outcome for participant *i* under the intervention and $Y_i(0)$ denotes the counterfactual outcome for participant *i* under the control. Secondary comparisons will be on the relative scale.

For all endpoints, we will test the **null hypothesis** of no change in outcomes due to the intervention with a two-sided test at the 5% significance level. We will also report point estimates and 95% confidence intervals for each effect measure and the arm-specific average outcomes. Standard error estimation will be with the estimated influence curve, and statistical inference will follow from the Central Limit Theorem (i.e., using the standard normal distribution).[4]

For all endpoints, we will also examine the intervention effect within subgroups defined by sex, age group (<60 years vs. ≥60 years), country, baseline HTN severity, and HIV status. In subgroups with fewer than 51 participants, we will conduct unadjusted analyses.

2.2 Evaluate the effect on hypertension control at week 24 and 48

The primary endpoint is hypertension control at week 24, evaluated in all participants in a research visit at week 24 and defined as blood pressure < 140/90 mmHg on the average of the second and third measurements. Participants who miss their week 24 visit and, thus, are missing blood pressure measurement at week 24 will be assumed to be uncontrolled.

As described above, we will use TMLE with Adaptive Pre-specification to compare hypertension control between intervention and control arms and to test the null hypothesis that the intervention did not impact hypertension control. We will conduct sensitivity analyses using TMLE to adjust for differences in characteristics (e.g., arm, age sex, sex, country, baseline severity) between persons with measurements and persons with missing measurements. We may conduct an additional sensitivity analysis excluding participants with missing values.

A secondary endpoint is hypertension control at week 48, assessed in the same manner as hypertension control at week 24.

At week 24 and week 48, we will also evaluate the intervention impact on hypertension severity, defined as participants with grade 2 or higher hypertension. For the severity endpoint, our primary approach will be TMLE to adjust for differences between persons with and without measured outcomes, and the secondary analysis will exclude participants with missing outcomes.

2.3 Evaluate the effect on retention in care at week 24 and 48

Using an analogous approach, we will formally compare retention in care by trial arm at week 24 and at week 48. We define retention in care as not late for the most recent scheduled hypertension care appointment by 30 days or more. This endpoint is not subject to missingness. Within TMLE and Adaptive Pre-specification, we will use the same candidate adjustment variables and approaches as the primary endpoint. Sensitivity analyses will be unadjusted.

2.4 Evaluate the effect on average systolic blood pressure at week 24 and 48

We will assess the intervention effect on mean systolic blood pressure at week 24 and week 48. Mean systolic blood pressure will be assessed using the average of the second and third blood pressure measurement taken by research staff at these timepoints. We will use the same candidate adjustment variables and estimation approaches as the primary endpoint.

In the primary approach, we will assume persons without a blood pressure measurement at the timepoint of interest are out-of-care and their blood pressure has reverted back to baseline. In other words, we will impute missing systolic blood pressure measures with their baseline measure. In sensitivity analyses, we will control for incomplete ascertainment of blood pressure measures. Specifically, we will use TMLE to adjust for differences in characteristics between persons with measurements and persons without measurements. For comparison, we may also conduct a complete-case analysis (i.e., restricting to participants with known outcomes.)

2.5 Evaluate the effect on time to hypertension control

Using routine clinical data, we may evaluate the intervention effect on the time to hypertension control (<140/90 mmHg). To conduct such an analysis, we would use Kaplan-Meier to compare the cumulative incidence of attaining hypertension control over the first 24 weeks using routine clinical data and censoring at death.

2.6 Assess predictors of hypertension control and retention in care

We may evaluate the following predictors of hypertension control and of retention in care at 24 and 48 weeks: sex, age, country, baseline HTN severity (grade), baseline HIV status (assessed via testing and self-report), and chronic kidney disease (CKD) at baseline. In such analyses, TMLE would be used to obtain variable importance measures, capturing the amount of information that a given predictor provides after adjusting for the other predictors. Secondary analyses would be unadjusted (i.e., univariate) associations.

2.7 Implementation outcomes

We will report and describe intervention fidelity; delivery of hypertension care in both arms, as well as barriers and facilitators of hypertension care engagement. Our implementation outcomes include

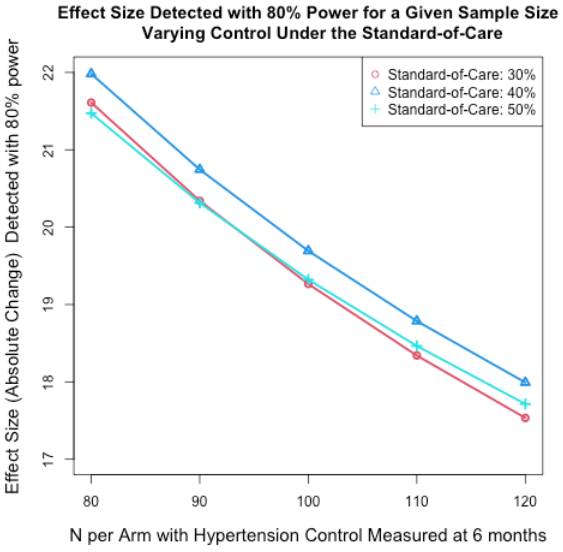
- Reach: completion of ≥1 post-baseline clinical visit
- Fidelity: clinician fidelity to hypertension guideline:
    - If BP controlled: continuing medications and scheduling in 12 weeks
    - If BP uncontrolled: increasing medications (if adherence good) or continuing medications (if adherence poor) and scheduling in 2-4 weeks
- Acceptability:
    - Description of barriers and facilitators reported by participants at baseline, week 24, and week 48 by trial arm (structured survey data)*
- Cost

*If participants were missing survey responses at week 48, we carried forward responses from week 24.

Appendix: Power calculations

Sample size and power calculations were based on standard formulas for a two-sample test of proportions and done with *power.prop.test* in *R*.[13] We expect these calculations to be conservative, because of the precision gained through stratified randomization and through covariate adjustment during the analysis.

We estimated 200 participants (~100/arm) would provide 80% power to detect at least a 20% absolute increase in hypertension control from 40% under the standard-of-care at 24 weeks (i.e., 6 months). Even with 20% fewer participants enrolled (from 100 to 80 participants/arm) and lower or higher control under the standard-of-care, these calculations suggest we would be well-powered to detect at least a 22% absolute increase in control.



References

1. Kwarisiima D, Atukunda M, Owaraganise A, et al. Hypertension control in integrated HIV and chronic disease clinics in Uganda in the SEARCH study. *BMC Public Health*. 2019;19(1):511. doi:10.1186/s12889-019-6838-6

2. Kwarisiima D, Kamya M, Owaraganise A, et al. High rates of viral suppression in adults and children with high CD4+ counts using a streamlined ART delivery model in the SEARCH trial in rural Uganda and Kenya. *J Int AIDS Soc*. 2017;Jul 21(20).

3. Havlir DV, Balzer LB, Charlebois ED, et al. HIV Testing and Treatment with the Use of a Community Health Approach in Rural Africa. *N Engl J Med*. 2019;381(3):219-229.

4. van der Laan M, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer; 2011.

5. Moore KL, van der Laan MJ. Covariate Adjustment in Randomized Trials with Binary Outcomes: Targeted Maximum Likelihood Estimation. *Stat Med*. 2009;28(1):39-64. doi:10.1002/sim.3445

6. Rosenblum M, van der Laan MJ. Simple, Efficient Estimators of Treatment Effects in Randomized Trials Using Generalized Linear Models to Leverage Baseline Variables. *Int J Biostat*. 2010;6(1):Article 13. doi:10.2202/1557-4679.1138

7. van der Laan MJ, Rose S. *Targeted Learning in Data Science*. Springer; 2018.

8. Colantuoni E, Rosenblum M. Leveraging Prognostic Baseline Variables to Gain Precision in Randomized Trials. *Stat Med*. 2015;34(18):2602-2617.

9. Balzer LB, van der Laan M, Ayieko J, et al. Two-Stage TMLE to Reduce Bias and Improve Efficiency in Cluster Randomized Trials. *Biostatistics*. 2021;kxab043. https://doi.org/10.1093/biostatistics/kxab043

10. Benkeser D, Díaz I, Luedtke A, Segal J, Scharfstein D, Rosenblum M. Improving precision and power in randomized trials for COVID-19 treatments using covariate adjustment, for binary, ordinal, and time-to-event outcomes. *Biometrics*. 2021;n/a(n/a):1-15. doi:10.1111/biom.13377

11. Balzer L, van der Laan MJ, Petersen M, SEARCH Collaboration. Adaptive Pre-specification in Randomized Trials With and Without Pair-Matching. Stat Med. 2016;35(10):4528-4545. doi:10.1002/sim.7023

12. Balzer LB, Cai E, Garraza LG, Amaranath P. Adaptive Selection of the Optimal Strategy to Improve Precision and Power in Randomized Trials. https://arxiv.org/abs/2210.17453v2. Published online 2023

13. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2020. http://www.R-project.org