STATISTICAL ANALYSIS PLAN (SAP)

Adolescent Wellness Visits in Tanzania (VITAA)

NCT number NCT05306938
Document Date 01/28/2026

Statistical Analysis Plan (SAP)

Title	Adolescent Wellness Visits (VITAA) in Tanzania, Aim 1
CRU/Department/Division/Center	Duke Global Health Institute
IRB Number	Pro00109406 (Duke) 21-1198 (UNC)
Investigators:	
Primary Investigator	Joseph Egger, Joy Noel Baumgartner
Collaborative Lead	Joy Noel Baumgartner
Co-authors (if known)	
Analysis Biostatistician(s)	David Arthur
Biostatistics Supervisor	John Gallis
Lead Biostatistician	John Gallis
Subject Matter Expert	Joe Egger
Original Creation Date	11/17/2021
Version Date	01/28/2026
Project Folder Location	PROJECT_DGHI_VITAA_Baumgartner\ \Projects\0 - Main Trial Paper\Docs
Project Goal(s)	Manuscript
Submission Deadline(s)	
clinicaltrials.gov	https://clinicaltrials.gov/ct2/show/NCT05306938
NIH Reporter	https://reporter.nih.gov/search/eJmGW5dWEkWkTjIqfsnGRA/project-details/10433988

Investigator Agreement

☐ All statistical analyses included in an abstract or manuscript should reflect the work of the biostatistician(s) listed on this SAP. No changes or additional analyses should be made to the results or findings without discussing with the project biostatistician(s).

☐ All biostatisticians on this SAP should be given sufficient time to review the full presentation, abstract, manuscript, or grant and be included as co-authors on any abstract or manuscript resulting from the analyses.

☐ If substantial additional analysis is necessary or the aims of the project change, a new SAP will need to be developed.

☐I have reviewed the SAP and understand that any changes must be documented.

	Acknowledged by: Joy Noel Baumgartner *Date:* January 28, 2026	
Activity Log	2/22/2022 – John updating based on newest IRB protocol 9/30/2022 – Kate adding in decision for new power calculation. 10/06/2022 – John updating details 11/30/2023 – More updates 01/31/2024 – Accepting tracked changes and making brief updates to the stats analysis section. 06/15/2025 – Updates incorporating changes discussed with study team based on example main results report 01/28/2026 – Minor revisions, remove comments and tracked changes	
Acronyms	*AWV*	*Adolescent Wellness Visits*
	CCR	*Covariate Constrained Randomization*
	CONSORT	*Consolidated Standards of Reporting Trials*
	cRCT	*Cluster Randomized Controlled Trial*
	CV	*Coefficient of Variation*
	FGD	*Focus Group Discussions*
	GEE	*Generalized Estimating Equations*
	HR	*Hazard Ratio*
	HTC	*HIV testing and counseling*
	ICC	*Intraclass Correlation Coefficient*
	IRB	*Institutional Review Board*
	KCMR	*Kauermann and Carroll Martingale Residual*
	KM	*Kaplan-Meier*
	LMICs	*Low and Middle-Income Countries*
	MD	*Mancl-DeRouen*
	NIH	*National Institutes of Health*
	PID	*Participant Identification Number*
	REDCap	*Research Electronic Data Capture*
	SRH	*Sexual and Reproductive Health*

Contents

1 Study Overview

Low utilization of health services among adolescents is an entrenched health system problem yet there is no routine adolescent health check-up platform in many low and middle-income countries (LMICs) on which to build. Adolescents are less likely to access HIV testing and counseling (HTC) than adults, while they shoulder a disproportionate burden of multiple poor sexual and reproductive health (SRH) outcomes, particularly in sub- Saharan Africa. A health system paradigm shift is needed for adolescent health services to move from a curative focus to a prevention-focused system that facilitates more regular interaction.

This project evaluates the impact of clinic-based 'Adolescent Wellness Visits' (AWV) coordinated between primary schools and health facilities that offer a package of evidence-based youth-friendly services on HIV testing uptake in Dar es Salaam and Pwani regions in Tanzania. About 900 adolescents will be enrolled in the cluster RCT from 20 primary schools, and a total of 250 adult stakeholders will be enrolled in focus group discussions (FGDs) and in-depth interviews. Procedures (methods): Adolescent data will include surveys with self-report measures and reviews of their medical records (using REDCap). The adult data will be in the form of translated and transcribed FGD and interview transcripts as Word documents. Regression models will be fitted to data to estimate the effect of the intervention on HIV testing and counseling (HTC) and contraceptive uptake at two years post- randomization. The study poses minimal risk to subjects.

This real-world pragmatic implementation trial will evaluate the impact of AWVs on multiple health service use outcomes for adolescents in Tanzania. The primary outcome is having any HTC up to two years post-primary school (~ages 14-16). In addition to contributing to knowing one's HIV status, HTC is a great indicator for preventive service use since it is applicable for all adolescents in Tanzania. Secondary outcomes for subpopulations include contraceptive use, linkages to HIV care and treatment, and treatment for health problems identified during screening.

1.1 Trial Design

This study is designed as a cluster randomized controlled trial (cRCT) with 20 Tanzanian primary schools (1 school=1 cluster) to evaluate the impact of Adolescent Wellness Visits (AWV) administered in the last year of primary school (Standard 7; mean age 13) on HIV testing and counseling (HTC) uptake among adolescents up to 2.5 years post-primary school (median age 15-16).

1.2 Study Aims

This SAP covers Aim 1 of the trial, which is to assess the impact of Adolescent Wellness Visits on HIV Testing and Counseling uptake among adolescents up to two years post-primary school utilizing a cluster randomized controlled trial.

1.3 Study Hypotheses

1.3.1 Primary Hypothesis 1

We hypothesize that the rate of uptake of HTC (**exclusive** of the Adolescent Wellness Visit) will be higher in the AWV intervention arm compared to control.

1.3.2 Primary Hypothesis 2

We hypothesize that the proportion of adolescents self-reporting HTC (**inclusive** of the Adolescent Wellness Visit) will be higher in the AWV intervention arm, compared to control.

1.3.3 Secondary Hypothesis

We hypothesize that reported contraceptive use at most recent sex will be higher in the intervention arm compared to control.

2 Study Population

2.1 Defining the Population under study

This project will be conducted in two sites in Tanzania:xxx District in Dar es Salaam region (urban) and xxx District in the Coastal region (rural/peri-urban). District health and education officers have confirmed support for allowing us to administer the AWVs within their catchment areas. Final identification of a list of matched school-clinic sites will be in collaboration with government officials prior to randomization. District enrollment data from schools was collected in 2019 for Standard 6 which indicates varying class sizes and this will be taken into account for constrained randomization. For xxx district, Dar es Salaam region, there were 10,645 Standard 6 students across 46 schools which averages 231 students per class cohort. For xxx district, Pwani region, there were 2,988 Standard 6 students across 124 schools, averaging 24 students per class cohort. Randomization occurs at the school level, stratified by region and healthcare facility type.

At each selected school, the entire Standard 7 class (stream) would be eligible to participate. While the mean age of Standard 7 students is 13 years old, we will have a minimum participant age at enrollment of 10 years and a maximum of 17 years. We plan to enroll about 900 participants across 20 primary schools.

2.2 Inclusion Criteria

- Currently enrolled and attending the study primary schools in Standard 7
- At enrollment, the minimum age is 10 and the maximum age is 17. (Per Tanzanian law, adolescents age 10 and older may legally consent to HIV testing)

2.3 Exclusion Criteria

- N/A

2.4 Power & Sample Size calculations

We assumed that the difference in proportion with HTC uptake between arms is 0.20, with 20% in control having HTC uptake and 40% in intervention having HTC uptake. In addition, we conservatively assume an ICC of 0.05, an alpha of 0.05, and a coefficient of variation of cluster sizes of 0.65. With these assumptions we will have 95% power to detect this effect with 10 clusters per arm (i.e., 20 clusters total) and 1200 participants total at the beginning of the trial (accounting for 20% dropout). Even if we lose 2 clusters per arm, we have 89% power. See ***Table 1*** below.

As we began the process of collecting data, it became evident that we were going to be able to recruit 900 participants total. With this new information, we retained the same assumptions as before: the difference in proportion with HTC uptake between arms is .20, with 20% in control having HTC uptake and 40% in intervention having HTC uptake, ICC = 0.05, alpha = 0.05, and the coefficient of variation of cluster sizes = 0.65. However, we altered our power calculation to have 900 participants at the beginning of the trial (accounting for 20% dropout) with 10 clusters per arm as before. With these updates, we will have 92% power to detect this effect and in the case of 2 clusters dropping out per arm, we will still have 86% power to detect the 0.2 proportion difference. See ***Table 2*** below.

Table 1. Power Calculation (original)

Power	***Clusters per arm***	ICC	Participants per cluster	Participants per cluster (inflated 20%)	Total number of participants	Total number of participants (inflated 20%)
Assuming 60 students per school at the beginning of the study						
0.97	***12***	0.05	48	60	1152	1440
0.96	***11***	0.05	48	60	1056	1320
0.95	***10***	0.05	48	60	960	1200
0.92	***9***	0.05	48	60	864	1080
0.89	***8***	0.05	48	60	768	960
0.85	***7***	0.05	48	60	672	840
0.79	***6***	0.05	48	60	576	720
0.71	***5***	0.05	48	60	480	600

Table 2. Updated Effect Size Detection with 900 students

Power	***Clusters per arm***	ICC	Participants per cluster	Participants per cluster analyzed (20% dropout)	Total number of participants	Total number of participants analyzed (20% dropout)
Assuming 45 students per school at the beginning of the study						
0.99	***10***	0.02	45	36	900	720
0.97	***10***	0.03	45	36	900	720
0.92	***10***	0.05	45	36	900	720

0.97	***8***	0.02	45	36	720	576
0.93	***8***	0.03	45	36	720	576
0.86	***8***	0.05	45	36	720	576

A.

2.5 Randomization

Given the relatively small number of randomization units (i.e., 20 school-clinic pairs), covariate constrained randomization (CCR) will be implemented to prevent severe imbalance in baseline covariates between intervention and control arms.[1] In addition, CCR can account for both stratification and covariate balance simultaneously, which is suitable for VITAA. Randomization was stratified by district and facility type and constrained on total class size and stream size (i.e., the actual number of students enrolled in the school). Attempts will be made to blind Joe Egger and the primary statistician (i.e., David) for the primary analysis. Specifically, David will create a Stata program that keeps only variables needed for the primary analysis. This will include removal of any 'AWV' related variables that could unblind the statistician. John Gallis will then create a separate Stata program file that will randomly re-allocate the binary arm designation of trial participants (i.e., the original 1/0 could be kept at 1/0 or changed to 0/1). John will then keep a record of this allocation decision but will not reveal this decision to David or Joe throughout the analysis. All data files used for blinded analyses will be kept in a separate Box folder labeled "VITAA_Unblinded", which John Gallis will only have access. All other secondary analyses will be blinded where possible. For example, the primary outcome will be used to estimate the effect of the treatment, both including and excluding the AWV visit baseline information. For the analysis including the AWV visit, the primary statistician will need to be unblinded because it will require AWV-related data.

The balance between arms on the constrained variables is shown in **Table 3**.

Table 3. Balance between arms on the constrained variables

	Arm 1 (N = 10)	Arm 2 (N = 10)
Total class size		
Mean (SD)	107.00 (90.62)	106.40 (69.46)
Class or stream size		
Mean (SD)	60.60 (19.43)	61.60 (21.44)
Facility Type		
Dispensary	8 (80.0%)	8 (80.0%)
Health Centre	2 (20.0%)	2 (20.0%)
District		
xxx	5 (50.0%)	5 (50.0%)
xxx	5 (50.0%)	5 (50.0%)

3 Outcomes, Exposures, and Additional Variables of Interest

3.1 Time of outcome assessment and visit windows

All study participants (intervention and control) will be interviewed at 4 timepoints: baseline and three follow-up time points (after the end of Standard 7).

Visit windows will be +/- 90 days of exact time point, to allow for difficulty in reaching study participants.

3.2 Primary Outcome(s)

Outcome	**Description and target population**	**Variable**	**Specifications**	***Baseline Survey + AWV Medical Records***	**Follow-Up 1**	**Follow-Up 2**	**Endline**
Number of adolescents who self-reported HIV Testing and Counseling (HTC),exclusive of the Adolescent Wellness Visit	Time to first self-reported HIV testing and counseling event, exclusive of the Adolescent Wellness Visit, assessed among all participants who are HIV negative at baseline	hiv_test_days	continuous: # of days from baseline interview to first reported HIV test NA if no test reported		X	X	X
Number of adolescents who self-reported HIV Testing and Counseling (HTC), – inclusive of the Adolescent Wellness Visit	Self-reported HIV testing and counseling, inclusive of the Adolescent Wellness Visit, assessed among all participants who are HIV negative at baseline	tested	0: No 1: Yes	X	X	X	X

3.3 Secondary Outcome(s)

				Time points variable is measured at			
Outcome	**Description and target population**	**Variable**	**Specifications**	***Baseline Survey + AWV Medical Records***	**Follow-Up 1**	**Follow-Up 2**	**Endline**
Number who reported contraceptive use at most recent sex, among participants who reported having sex during the study	Assessed among those who are sexually active (reported having sex during the study)	contracep_last_sex	0: No 1: Yes	X	X	X	X

4 Statistical Analysis Plan

Statistical analysis will be conducted according to the CONSORT guidelines. A flow chart will show the participation of both intervention and control arms in terms of eligibility screening, recruitment, and follow-up status. All analyses will focus on comparison of the two arms. The timeline cluster tool will be used to identify potential selection bias if needed.

The overall trial success is defined as achieving statistical significance for both Primary Outcome 1 and Primary Outcome 2 simultaneously. Because the rejection of both null hypotheses is required to claim study success, no adjustment for multiplicity is necessary. Each primary hypothesis will be tested at the full two-sided significance level of 0.05.

4.1 Demographic and Clinical Characteristics ("Table 1")

We will summarize individuals' baseline characteristics by study arm, using means, standard deviations, medians, etc. for continuous variables and frequency and percentage for categorical variables. We will also report school-level variables summarized by intervention arm. Any individual or school characteristic differential by intervention arm at baseline (based on $p<0.10$) will be candidate variables to be included in a sensitivity analysis adjusted for these variables.

4.1.1 Data Structure for Primary (Survival) analysis

Data will need to be in wide format with a unique observation for each PID.

- **PID**
- **School ID (cluster variable)**
- **Event**: HTC (0/1)
- **Origin date**: date of baseline survey (check to make sure that AWV visit and BL survey date are the same day for intervention date – this should be done outside of main analysis to maintain blinding).
- **Censored date:** date participant is censored from study. Takes value of endpoint if participant is not censored for any other reason (alive, no HTC, no ltfup).
- **Time** (days): coded as censored date minus origin date, as integer.
- **Covariates**

4.2 Analysis Plan for Aim 1 Primary Outcome 1

4.2.1 Analysis Plan for Aim 1 Primary Outcome 1 (Survival)

Analysis on aim 1 will test the primary hypothesis that randomization to the intervention arm will result in a lower rate of 'survival' to first HIV test, compared to the control arm during the first approximately 2.5 years of follow-up time.

The origin time will be set as the baseline interview date for all participants. Participants will be censored for the following reasons:

- Administratively censored at their endline interview date (survival to the end of the study without experiencing the event).
- Lost to follow-up. Interval censored. Participant is censored at their last interview date.
- Death (competing event). Participant is censored on date of death.

The following is the primary outcome event of interest:

- Reported first HIV test. This is either exact or approximate. If date is either exact or approximate, participant is censored on that date.

Time will be recorded as time of survival in days since baseline interview data. The outcome will be coded as a 0/1 where 1= whether they have the event (HTC) during follow-up.

We will use intention-to-treat to analyze the data using a marginal Cox proportional hazards model framework with a correction for small cluster sample size.[2] Per recommendations, we will use a MD correction with a Wald t-test for inference if the CV of cluster sizes is < 0.4, or a KCMR correction if the CV of cluster sizes is greater than 0.4.[2] We will decide based on the data whether we need to also make a correction for competing risks. Participants will be considered censored at the point of the last survey at which they were non-contactable, including death. Importantly, for primary outcome 1 (HTC, exclusive of the Adolescent Wellness Visit), participants in the AWV arm who receive an HIV test during the AWV will not be counted as having the event of interest on the baseline interview date but are eligible to be censored or have the event on any day following the baseline interview date. A secondary analysis can estimate the 2-year cumulative incidence of HTC separately for inclusion or exclusion of these participants (those that received an HIV test at AWV).

We will first plot an unadjusted Kaplan-Meier (KM) curve. Then, we will fit a marginal Cox proportional hazards model with a GEE (independence working correlation matix) to estimate the hazard ratio of HTC comparing the AWV arm to the control arm. We will use the CoxBcv command in R to fit an appropriate small-sample correction in order to obtain valid type I error with fewer than 30 clusters.[2] All variables used in the constrained randomization will be included in the primary statistical model. We will also include in the primary statistical model any additional baseline covariates (Table 1) that predict missing outcome data or which are imbalanced across arms at baseline. To ensure the rigor and validity of the effect estimates, as a sensitivity analysis, we will repeat the analysis using a model that includes only the design variables (those used in the constrained randomization) as adjustment covariates.

Due to the inherent selection bias of hazard ratios over long follow up times, we will report both an average HR over the full follow-up, as well as an HR for the period from baseline interview to first follow-up interview. Any differences in this HR over time will be due either to the inherent selection bias, whereby susceptibles are consumed differentially by treatment arm leading to an apples/oranges comparison, or due to a change in the underlying hazard ratio over time. To investigate this further, we may consider constructing weighted survival curves, weighting by changes in the population structure over time.

We will also evaluate the modifying effect of sex and age on the intervention effect on HTC. For this moderation analysis, we will separately include interaction terms between intervention and our pre-defined potential modifiers, evaluating the changes in the effect on HTC.

4.2.1.1 Data sparsity issue for KM curves

For logistical reasons, particularly the fact that some participants were challenging to reach, there was an extended interview window at each timepoint. This resulted in data where some participants were interviewed relatively early for their baseline interview, and then later than average at follow-up, leading to a longer-than-average time under study. These participants fall into the tails of the survival time distribution. Due to the small number of participants eligible for the event at these later survival times, we face a precision issue with estimating the survival function (S(t)) during this period, where a single event can result in a precipitous drop in the KM curve.

To address this issue, we will use the method proposed by Gebski et al[3] (2018) to determine the minimum number of subjects remaining at risk, beyond which the KM curve should be curtailed. Following their "Criterion 2", we will calculate a minimum acceptable number of subjects still at

risk by comparing the size of the decrease in (S(t)) if an extra event occurs with the variability of the survival estimate had all subjects been followed to that time, using a confidence interval approach. Specifically, we will curtail the KM curve at the time *t* where one extra event observed just after time *t* would decrease the estimated survival probability to below its full information one-sided 95% confidence boundary at time *t*.

4.3 Analysis Plan for Aim 1 Primary Outcome 2

Unlike Primary Outcome 1, this analysis does count tests received during the AWV as the event of interest. Therefore, a large "spike" in events is expected for the AWV arm shortly after the baseline interview, meaning the proportional hazards assumption required for Cox regression likely will not hold. For that reason we will analyze this as a binary outcome: whether an HIV test was reported during the study period or not).

To estimate the risk ratio comparing HTC (inclusive of the AWV) between the two arms, a modified Poisson generalized estimating equations (GEE) model will be used to estimate the intervention effect at the final timepoint.[4] The model will include an exchangeable correlation structure to account for the correlation of outcomes among adolescents within the same cluster. We will also use Kauermann-Carroll bias-corrected variances to account for the relatively small number of clusters in our study.[5] All variables used in the constrained randomization will be included as adjustment covariates in the statistical model.

4.4 Analysis Plan for Aim 1 Secondary Outcomes

To estimate the risk ratio comparing contraceptive use between the two arms, a modified Poisson generalized estimating equations (GEE) model will be used to estimate the intervention effect at the final timepoint.[4] The model will include an exchangeable correlation structure to account for the correlation of outcomes among adolescents within the same cluster. We will also use Kauermann-Carroll bias-corrected variances to account for the relatively small number of clusters in our study.[5] All variables used in the constrained randomization will be included as adjustment covariates in the statistical model .

5 References

1. Gallis JA, Li F, Yu H, Turner EL. cvcrand and cptest: Commands for efficient design and analysis of cluster randomized trials using constrained randomization and permutation tests. *Stata Journal* 2018; **18**(2): 357-78.
2. Wang X, Turner EL, Li F. Improving sandwich variance estimation for marginal Cox analysis of cluster randomized trials. *Biometrical Journal* 2023; **65**(3): 2200113.
3. Gebski V, Gares V, Gibbs E, Byth K. Data maturity and follow-up in time-to-event analyses. *Int J Epidemiol* 2018; **47**(3): 850-9.
4. Zou G. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol* 2004; **159**(7): 702-6.
5. Kauermann G, Carroll RJ. A note on the efficiency of sandwich covariance matrix estimation. *J Am Stat Assoc* 2001; **96**(456): 1387-96.