Sexually transmitted infections in CHIEDZA (STICH)

Statistical analysis plan

Version 1.0
Version 1.0 started 28 April 2022
Clinicaltrials.gov identifier: NCT 03719521
MRCZ reference: A/2387

Trial statistician: Victoria Simms

Signature …… Date……..18 Aug 2022…..

Chief investigator: Chido Dziva Chikwari

Signature …… Date…18 August 2022……

Chair of Trial Steering Committee: David Ross

Signature …… …… Date …18th August 2022

Amendments to SAP after signoff of version 1.0

Date	SAP version	Protocol version and date	Reason for change
	V1.0	V4.0, 30 April 2021	Initial version

Contents

1.1. Investigators

Principal investigator
Name: Chido Dziva Chikwari
Address: Biomedical Research and Training Institute. 8 Ross Avenue, Harare, Zimbabwe
Email: Chido.Dzivachikwari@lshtm.ac.uk

Trial statistician
Name: Victoria Simms
Address: Biomedical Research and Training Institute. 8 Ross Avenue, Harare, Zimbabwe
Email: Victoria.simms@lshtm.ac.uk

Data manager
Name: Tsitsi Bandason
Address: Biomedical Research and Training Institute. 8 Ross Avenue, Harare, Zimbabwe
Email: tbandason@brti.co.zw

Study coordinator
Name: Ethel Dauya
Address: Biomedical Research and Training Institute. 8 Ross Avenue, Harare, Zimbabwe
Email: edauya@brti.co.zw

Senior statistician
Name: Helen Weiss
Address: London School of Hygiene & Tropical Medicine. Keppel Street, London WC1E 7HT
Email: helen.weiss@lshtm.ac.uk

Senior statistician
Name: Richard Hayes
Address: London School of Hygiene & Tropical Medicine. Keppel Street, London WC1E 7HT
Email: Richard.hayes@lshtm.ac.uk

1.2. Contents of this SAP

This SAP refers solely to the main analysis of the STICH 16-arm cluster randomised controlled trial and to its primary, secondary and exploratory outcomes. The analysis of the 24-cluster CHIEDZA trial within which the STICH trial is embedded will be described in a separate CHIEDZA analytical plan. The STICH process evaluation, cost and cost-effectiveness analysis, and the analysis of the mixed-methods study of STI testing knowledge, attitudes and practice and acceptability of vaginal swab will also be described separately.

2. Description of the trial

2.1. Background

CHIEDZA (Community based interventions to improve HIV outcomes in youth: a cluster randomised trial in Zimbabwe) is a cluster-randomised trial of a youth-friendly service. Its aim is to determine the impact of an integrated community-based package of HIV services incorporating HIV testing, linkage to care and ongoing adherence support, combined with sexual and reproductive health (SRH) services and general health counselling for 16-24year-olds on population level HIV viral load among HIV-positive adolescents aged 18-24 years in a high HIV prevalence setting.

The CHIEDZA intervention is a community-based package of services that includes: HIV testing and counselling, delivery of antiretroviral therapy, adherence support groups, mHealth related to SRH messaging, provision of condoms, menstrual hygiene management, contraception and treatment of sexually transmitted infections, referral for voluntary medical male circumcision, risk reduction counselling and general health information and counselling. The intervention will be implemented over a 30-month period in each cluster. The intervention will be implemented in 3 provinces, and 4 clusters in each province, each with a population of approximately 2500 16–24-year-olds. These will be compared to 4 control clusters in each province.

STICH (Sexually Transmitted Infections in CHIEDZA) is an intervention consisting of outreach, promotion and mobilisation strategies applied to STI services, with unselected testing for chlamydia and gonorrhoea using urine tests (males and females), and self-collected vaginal swab testing for trichomonas (females), with treatment offered for those positive and their partners. STICH services are delivered at CHIEDZA centres. The intervention will be implemented in 2 provinces (Harare and Bulawayo) and 4 clusters in each province, for a 12-month period, staggered by 3 months. These intervention clusters will be compared to 4 control clusters in each of the 2 provinces. STICH outcomes will be measured in a cross-sectional survey which is embedded within the CHIEDZA outcomes prevalence survey.

2.2. Objectives

2.2.1. Primary objective

The primary objective of this study is to measure the impact of unselected STI testing and treatment of young people on population-level prevalence of any of *Chlamydia trachomatis*, *Neisseria gonorrhoeae* and *Trichomonas vaginalis*.

2.2.2. Secondary objectives

The secondary objectives of the study are to:

- Measure the impact of STI testing and treatment of young people on population-level prevalence in 18-24 year olds of
    - *Chlamydia trachomatis*
    - *Neisseria gonorrhoeae*
    - *Trichomonas vaginalis*

- Determine uptake, prevalence, and yield of STI testing
- Determine uptake of partner notification
- Describe the acceptability of self-collected vaginal swabs
- Determine factors predicting acceptance of STI testing

- Assess knowledge, attitudes, and practices towards STIs, STI testing and partner notification using a mixed-methods study

Prevalence is defined as the proportion of test results that are positive. Yield is defined as the proportion of positive test results out of the number of clients who were eligible for testing.

2.3. Trial design including blinding

A cluster randomised parallel arm trial design will be used to address the objectives of the study. The study will be conducted in two provinces of Zimbabwe (Harare and Bulawayo). A total of 16 study clusters, stratified by province, will be randomised in a 1:1 allocation ratio to one of the following arms:

i. Control Arm - Existing SRH services
ii. Intervention Arm - Community-based intervention including both the CHIEDZA and STICH interventions

Random allocation to the CHIEDZA intervention took place at public randomisation ceremonies in each province. The STICH intervention will be allocated to the same clusters randomly allocated to the CHIEDZA intervention.

Outcomes will be measured at population level with a community cross-sectional survey. The end-line survey will recruit a random sample of 300 participants aged 18-24 years in each of the 16 clusters (8 intervention and 8 control clusters) (total 4,800 participants). Data collection will take place between October 2021 and March 2022. The sampling strategy is explained in section 2.5.

STICH is an “open label” trial. Data collectors for the prevalence survey will not be blinded to trial arm. The trial statistician will also be unblinded to trial arm.

A final cleaned dataset will be produced after the last survey visit has been completed. Distributional assumptions will be tested, exclusions will be defined, and the analysis plan will be finalised. A non-blinded analysis will then be conducted by the trial statistician. Semi-blinded trial results, with trial arms labelled A and B, will be discussed, and interpreted at an investigators´ meeting. An envelope containing the treatment allocation will be opened after this discussion at the investigators´ meeting.

2.4. Eligibility for the intervention

Inclusion criteria

- Lives within one of the STICH intervention clusters
- Aged 16-24 years
- Female (for Trichomonas test only)
- Attends CHIEDZA service between 21 September 2020 and 30 September 2021 (Harare) or 1 January 2021 and 31 December 2021 (Bulawayo), the times when the STICH intervention is operational

Exclusion criteria

- Has had an STI test within the previous 3 months, unless they are symptomatic and request an STI test

2.5. Eligibility for the survey

For the prevalence survey all 16 clusters were mapped using OpenStreetMap and the road networks were divided into sections of approximately 200 metres in length. Sections were each allocated an ID number, and simple random sampling was used to select 40 street sections and a backup list of 40 extra sections, in each cluster. The list of selected street sections was sent to the study coordinator. The study coordinator arranged for the survey team to survey all selected street sections. If the first list was completed before 700 participants had been enrolled, the team moved on to the backup list.

The entire survey aimed to enrol 700 participants per cluster, to meet the requirements of the CHIEDZA trial. The sample size for STICH is 300 per cluster. Therefore, data collection days of the survey were randomly allocated to STICH or non-STICH, stratified by weekday/Saturday, with 55% of days allocated to STICH. All participants recruited on STICH days were asked to provide a urine sample which will be tested for 5 STIs using Seegene (Allplex™ STI Essential Assay). Vaginal swab samples were not used in the prevalence survey. The 5 STIs are *C. trachomatis*, *N. gonorrhoeae*, *T. vaginalis*, *Ureaplasma urealyticum* and *Mycoplasma genitalium*. 55% of days were allocated to STICH because it was estimated that data collection would be slower on STICH days (due to the urine sampling) and so 55% of days are required to obtain 43% (300/700) of the participants.

Data collectors visited every dwelling in the allocated street sections and enumerated household members. If any household members were aged 18-24 years, they approached these potential participants for consent. If eligible participants were not at home the data collectors made up to 3 return visits on three different days.

Household members aged 18-24 were shown a video which explains the survey procedures and gave consent if they agreed to participate. The video includes details of STICH including the need for a urine sample and the tests to be performed. Paper consent forms were used in case of video failure, and for a small number of participants as a control group for a randomised controlled trial of consent methods.

Inclusion criteria

- Aged 18-24 years
- Consent to participate
- Surveyed on a day allocated to STICH sampling

Exclusion criteria

- Unable or willing to provide informed consent
- Unable or unwilling to provide a urine sample

2.6. Outcome variables

2.6.1. Primary outcomes

The primary outcome is the proportion of participants with a positive test result for any of *Chlamydia trachomatis, Neisseria gonorrhoeae* and/or *Trichomonas vaginalis,* compared between trial arms. Participants will provide a urine test which will be tested for GC, CT and TV. Indeterminate test results will be excluded from the primary outcome.

2.6.2. Secondary outcomes

The secondary outcomes are the proportion of participants with each of the following STIs, compared between trial arms:

- *Chlamydia trachomatis*
- *Neisseria gonorrhoeae*
- *Trichomonas vaginalis*

2.6.3. Survey outcomes

The assay will also test for *Ureaplasma urealyticum* and *Mycoplasma genitalium*. The prevalence of *Ureaplasma urealyticum* and *Mycoplasma genitalium* will be reported overall, but not compared by trial arm.

All participants will be asked to give a dried blood spot (DBS) sample which will be tested for HIV and for herpes simplex (HSV-2). The prevalence of HSV-2 will be also reported overall as an exploratory STICH study outcome but will not be compared by trial arm. HIV status will be used for subgroup analysis (section 4.6).

2.7. Sample size considerations

The expected prevalence of the primary outcome in the control arm is 17%, based on a pilot survey of 2331 young people in CHIEDZA in 2019. The coefficient of variation in the outcome is conservatively estimated at 0.25-0.3. A sample size of 300 per cluster in 16 clusters gives 80% power to detect an absolute decrease in the outcome of 7% (to 10% prevalence) and 90% power to detect an absolute decrease of 8% (to 9% prevalence) at a coefficient of variation of 0.3 (Table 1). The trial is not powered to detect an effect on secondary outcomes. However, the minimum detectable difference for the CT and NG outcomes is shown in Tables 2 and 3.

Table 1: estimated minimum detectable difference for primary outcome

Cluster size	Clusters	Expected outcome in control arm	Power	Coefficient of variation	Maximum prevalence in intervention arm that will show an effect as being unlikely ($p<0.05$) to be due to chance
300	16	17%	80%	0.25	11%
300	16	17%	80%	0.3	10%
300	16	17%	90%	0.25	10%
300	16	17%	90%	0.3	9%

Table 2: estimated minimum detectable difference for CT outcome

Cluster size	Clusters	Expected outcome in control arm	Power	Coefficient of variation	Maximum prevalence in intervention arm that will show an effect as being unlikely (p<0.05) to be due to chance
300	16	14%	80%	0.25	9%
300	16	14%	80%	0.3	8%
300	16	14%	90%	0.25	8%
300	16	14%	90%	0.3	7%

Table 3: estimated minimum detectable difference for NG outcome

Cluster size	Clusters	Expected outcome in control arm	Power	Coefficient of variation	Maximum prevalence in intervention arm that will show an effect as being unlikely (p<0.05) to be due to chance
300	16	3%	80%	0.25	1%
300	16	3%	80%	0.3	1%
300	16	3%	90%	0.25	1%
300	16	3%	90%	0.3	1%

3. Data analysis plan – data description

3.1. Recruitment and representativeness of survey participants

A CONSORT flow chart for cluster-randomised trials will be constructed showing the number of clusters eligible and the number allocated to each arm. It will also show the enumerated population on STICH recruitment days, number of residents aged 18-24 years, number who were contacted, number agreeing to participate in the survey, and reasons for non-enrolment. It will also include the number of participants by trial arm who began the survey but did not complete it or who did not give a urine sample, with reasons if known.

3.2. Characteristics of the study population

Characteristics of participants will be described by trial arm (Table 1). As the trial has no pre-intervention measurement, only variables which cannot be affected by the STICH intervention will be reported here. Statistical tests for imbalance between trial arms will not be carried out because any difference is due to chance by definition if the randomisation has been conducted correctly. Demographic variables will be described by trial arm and overall: means and standard deviations/medians and interquartile ranges for continuous variables that are normally distributed/not normally distributed, and frequencies and proportions for binary or categorical variables. The list of variables we will describe is as follows: age, sex, relationship status, educational level attained, sexual debut.

3.3. Linkage

All STICH intervention clients who consented to it, and all outcome survey participants, were registered using a SIMPRINTS biometric system which will record their fingerprints and assign a unique ID. Fingerprint IDs from the prevalence survey will be linked to those from the implementation dataset, by province. This will enable comparison of self-reported STI testing to STICH records of STI testing as exploratory analysis (described below).

4. Data analysis – inferential analysis

4.1. Principles

CONSORT guidelines for cluster randomised trials will be followed. Cluster-level analyses will be used to adjust for between-cluster variability, as recommended for trials with fewer than 15 clusters per arm [1]. Descriptive analysis will be used to compare the cluster-level characteristics of the two arms, with adjustment for demographic variables that are imbalanced between arms and for province. Imbalance will be defined as a difference in distribution between arms likely to affect the results of the trial.

The main statistical analyses will follow the intention to treat (ITT) principle as much as possible. The trial statistician will be unblinded. The senior statisticians will remain partially blind (knowing only coded trial arm membership) until the primary analysis is complete. All participants from intervention and control clusters will be included whether or not they have participated in the CHIEDZA or STICH interventions.

4.2. Analysis of primary outcome

The risk of the three study STI infections (any of CT, NG or TV) for each cluster will be calculated and shown by province and arm. All participants with a valid STI urine test result will be included. The mean and SD of the log risk will be used to estimate the geometric mean and associated 95%CI for each arm of the study. Linear regression of the log mean risk on province and arm will be used to estimate the risk ratio and 95%CI (Table 2). The approximate variance for the mean risks will be obtained based on the residual mean square from a 2-way ANOVA on arm and province. A 95% CI for this will be calculated from the variance using a t-statistic. Significance tests will be two-sided with 5% level of significance.

4.3. Analysis of secondary outcomes

All three secondary outcomes are binary variables. Analysis for the secondary outcomes will follow the methods described in section 4.2 for the primary outcome.

4.4. Statistical considerations

4.4.1. Missing data

The number and proportion of participants with complete data for the primary and secondary outcomes will be reported by trial arm and overall. No multiple imputation is planned to correct for missing data in the primary and secondary outcomes, as the proportion of missing data is expected to be low given that there is no time period for loss to follow-up.

4.4.2. Model assumption checks

Model assumptions will be checked using standard methods.

4.5. Sensitivity analysis

The characteristics (age and sex) of trial participants will be compared with those of the enumerated population aged 18-24 on days allocated to STICH sampling. If there is evidence of response bias in those who participated in the survey, standardisation based on the enumeration data will be used on the primary and secondary outcomes to adjust for non-participation and obtain a result more representative of the enumerated population.

4.6. Planned subgroup analyses

No formal subgroup analyses are intended as the study is not powered to detect treatment effects for subgroups. However, as exploratory analysis, outcomes will be reported by sex (male and female), by HIV status (positive and negative), and among only those who report they have ever had sex. The definitions of these variables are shown below. For the sex and HIV status variables, effect modification will be examined. To do this, the cluster-level risk of STI infection will be calculated separately for the two groups (e.g. male and females), and then the mean log risk ratio will be compared between arms. An unpaired t-test on the cluster-level differences will be used to test for interaction.

4.7. Exploratory analysis

The STICH intervention is expected to have direct effects on clients who receive testing and treatment, and indirect effects on survey participants who do not attend the service but are exposed to a lower level of circulating STIs in the population as a result of the testing and treatment of their peers. As a form of exploratory analysis, we will assess the difference in primary and secondary outcomes between intervention arm participants who accessed STI testing through the STICH intervention compared to control arm participants. Access to the STICH intervention will be determined in two ways, i) via linkage of fingerprints collected in the intervention and the survey, ii) via self-report.

Men were not eligible for TV testing in the intervention, however, were tested for TV in the survey. As an exploratory analysis we will assess the difference between arms in the prevalence of CT and/or NG, by sex (in other words, a modification of the primary outcome that excludes TV).

4.8. Interim analysis

No interim analysis is planned in this study because no trial data is available prior to the endline survey.

4.9. Definitions of other variables

<u>Sex</u>
Participants are asked for their sex defined at birth and for their self-identified gender. For exploratory analysis by sex, sex defined at birth will be used.

<u>HIV status</u>

All participants will be asked to give a dried blood spot (DBS) sample which will be tested for HIV. Participants who refuse to give a DBS or whose test result is indeterminate will be excluded from exploratory analysis by HIV status.

Sexual debut

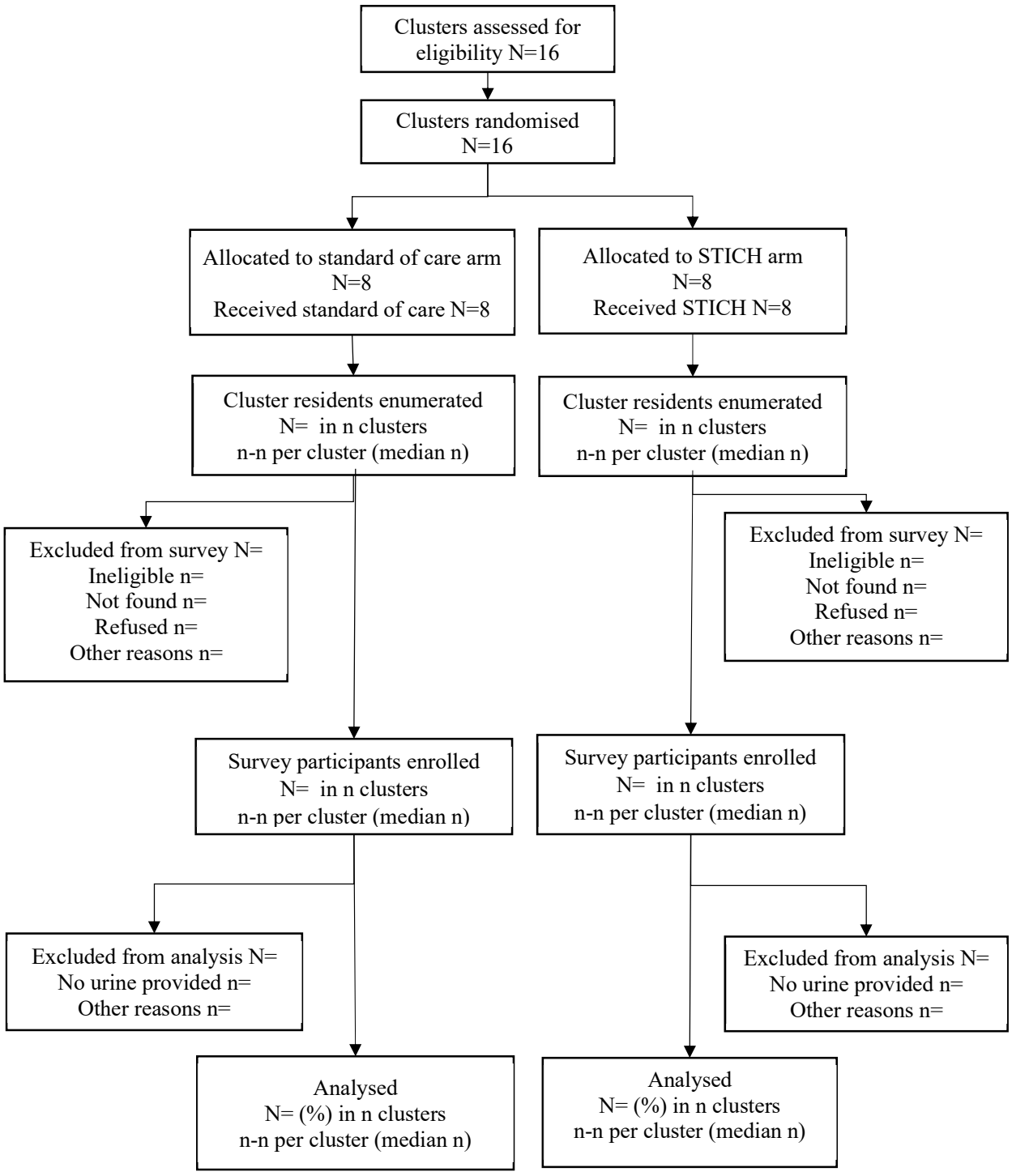
Participants will be asked whether they have ever had penetrative sexual intercourse, defined as follows: "*By this I mean when a man/husband/boyfriend/partner puts his penis inside a woman's vagina (vaginal sex), or inside a woman or a man's anus or backside (anal sex). Sex can be between two people of the same gender and can have occurred even if someone was not sure they wanted it*."

5. Software

Data management: for the survey, data will be collected using SURVEYCTO and uploaded to a server hosted at the Biomedical Research and Training Institute (BRTI) at 8 Ross Avenue, Zimbabwe. The data manager will extract data periodically as needed.

Statistical analysis: Stata 17.0 will be used for data description and for the main inferential analysis.

6. Figure 1: CONSORT Flow Diagram STICH trial



Any other reasons for exclusion that make up more than 5% of exclusions will be reported separately

Table 4: participant characteristics

Characteristic		CHIEDZA arm	Standard of care arm
Demographic			
Sex assigned at birth	Male		
	Female		
Self-identified gender	Male		
	Female		
Age	18-20		
	21-24		
Education level attained	Did not complete primary		
	Completed primary		
	Completed Form 4		
	Completed Form 6		
	Post-secondary		
Household monthly income	<USD 50		
	USD 50-100		
	USD 101-200		
	USD 201-500		
	>USD 500		
	Missing		
Main activity	Education		
	Formally employed		
	Informally employed		
	None of the above		
Marital status	Married or living together		
	Never married		
	Divorced, widowed or separated and currently unmarried		
Sexual debut	Ever had penetrative intercourse		
	Never had penetrative intercourse		

Table 5: Effect of the intervention on the primary outcome.

Outcome		STICH arm	Standard of care arm	Effect estimate	Test
		N=8	**N=8**		
Cluster-level proportion of participants positive for CT, NG or TV	Geometric mean prevalence (95% CI),			Risk ratio (95% CI)	Linear regression of log mean risk adjusting for strata and for variables imbalanced between arms

Table 6: effect of the intervention on secondary outcomes

Outcome		STICH arm	Standard of care arm	Effect estimate	Test
		N=8	**N=8**		
Cluster-level proportion of participants positive for CT	Geometric mean prevalence (95% CI),			Risk ratio (95% CI)	Linear regression of log mean risk adjusting for strata and for variables imbalanced between arms

Cluster-level proportion of participants positive for NG	Geometric mean prevalence (95% CI),			Risk ratio (95% CI)	Linear regression of log mean risk adjusting for strata and for variables imbalanced between arms
Cluster-level proportion of participants positive for TV	Geometric mean prevalence (95% CI),			Risk ratio (95% CI)	Linear regression of log mean risk adjusting for strata and for variables imbalanced between arms

7. References

1. Hayes, R.J. and L. Moulton, *Cluster Randomised Trials*. 2009, Boca Raton: Taylor & Francis.