Field Studies on the Feasibility of Interrupting the Transmission of Soil-transmitted Helminths (STH)
DeWorm3 Project

NCT03014167

January 2, 2024

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Department of International Health

January 2, 2024

Re: NCT03014167

Updated final SAP for ClinicalTrials: Field Studies on the Feasibility of Interrupting the Transmission of Soil-transmitted Helminths (STH) DeWorm3 Project

To Whom It May Concern:

Please find an updated SAP attached

Sincerely,

Judd L Walson, MD, MPH

Robert E. Black Chair, Department of International Health
Professor, Departments of International Health and Medicine (Infectious Disease)
Bloomberg School of Public Health
Johns Hopkins University

1. ADMINISTRATIVE INFORMATION

FIELD STUDIES ON THE FEASIBILITY OF INTERRUPTING THE TRANSMISSION OF SOIL-TRANSMITTED HELMINTHS (STH), NCT03014167

Current Document			
Version No.:	1.0		
Developed by:	Kristjana Ásbjörnsdóttir	**Date:**	04. Sep. 2023
Reviewed by:	Katherine Thomas and Sean Galagan, Unblinded Data Team	**Date:**	08. Dec. 2023
Effective Date:	28. Nov. 2023		

Approvals

Kristjana Á

18. Dec. 2023

Kristjana Ásbjörnsdóttir, Scientific Integrity Lead **Date**

Barbra Richardson

Digitally signed by Barbra Richardson
Date: 2024.01.03 07:43:11 -08'00'

Barbra Richardson, Study Statistician **Date**

02 Jan 2024

Judd Walson, Principal Investigator **Date**

2. INTRODUCTION

2.1 BACKGROUND AND RATIONALE

Soil-transmitted helminths (STH) are group of intestinal parasites (*Ascaris lumbricoides*, *Ancylostoma duodenale*, *Necator americanus* and *Trichuris trichiura)*. These parasites infect an estimated 1.45 billion people globally, resulting in the loss of almost 5 million disability adjusted life years (DALYs) annually [1, 2]. High to moderate intensity STH infections are associated with increased risk of malnutrition, iron-deficiency anemia and other adverse physical and cognitive morbidities, particularly in children. The current World Health Organization (WHO) strategy for controlling STH is based on mass drug administration (MDA) of albendazole or mebendazole to pre-school and school-age children, women of childbearing age (including pregnant women in the second and third trimesters and breastfeeding women) and adults in certain high-risk occupations such as agricultural laborers or miners, with the goal of eliminating STH-related morbidity [3-5]. While the current strategy of MDA for STH control has proven successful in controlling morbidity, the incidence of reinfection post-treatment is high and targeted MDA is unlikely to break transmission of STH in many settings [6]. This is in large part due to adult reservoirs of disease [7, 8]. Even if coverage of all targeted populations were optimized, the current STH strategy would likely need to be continued until significant economic development and universal water, sanitation and hygiene (WASH) access are able to interrupt transmission [9]. As a result, there is increased interest in alternative approaches to interrupt STH transmission using broadly administered MDA strategies targeting all age groups [10], which a recent systematic review and meta-analysis showed are more effective in reducing the prevalence of STH infection [11].

2.2. OBJECTIVES

DeWorm3 GENERAL OBJECTIVE: To determine the feasibility of interrupting the transmission of STH (*A. lumbricoides, A. duodenale, N. americanus* and *T. trichiura*) in focal geographic areas in Africa and Asia by expanding the population targeted and the frequency of delivery of MDA with albendazole.

2.2.1 PRIMARY OBJECTIVES

In focal geographic areas in India, Malawi and Benin, where LF programmes have delivered at least five rounds of MDA with albendazole (plus ivermectin or DEC);

Objective 1.1

To compare the prevalence of the predominant STH species *(A. lumbricoides, A. duodenale, N. americanus* or *T. trichiura)* as measured by quantitative PCR, 24 months after stopping MDA, between clusters randomized to receive twice-yearly community-wide MDA versus clusters randomized to receive standard of care pre-SAC and SAC targeted MDA.

Hypothesis: In settings where the prevalence and intensity of STH infections have been reduced through annual MDA of albendazole delivered by LF elimination programs, an additional three years of twice-yearly community-wide MDA with albendazole will result in a lower final cross-sectional prevalence of the predominant STH species driving local transmission as compared to a strategy of pre-SAC and SAC targeted MDA.

Objective 1.2

To determine whether the transmission of the predominant STH species *(A. lumbricoides, A. duodenale, N. americanus* or *T. trichiura)* can be interrupted using MDA with albendazole.

Transmission interruption will be assessed at the cluster level and is defined as reaching a weighted point prevalence of ≤2% of that species as measured by quantitative PCR 24 months after stopping MDA. The proportion of clusters randomized to twice-yearly community-wide MDA versus standard of care pre-SAC and SAC-targeted MDA that interrupt transmission will be compared.

Hypothesis: In settings where the prevalence and intensity of STH infections have been reduced through annual MDA of albendazole delivered by Lymphatic filariasis elimination programs, an additional three years of twice-yearly community-wide MDA with albendazole will be sufficient to interrupt transmission of the predominant STH species driving transmission at a particular study site, whereas MDA targeting pre-SAC and SAC will not achieve interruption of STH transmission.

2.2.2 SECONDARY OBJECTIVES

Objective 2.1

To compare the prevalence of all STH species *(A. lumbricoides*, *A. duodenale*, *N. americanus* and *T. trichiura*) in clusters receiving either twice-yearly community-wide MDA or standard of care pre-SAC and SAC targeted MDA.

Hypothesis: In settings where the prevalence and intensity of STH infections have been reduced through annual MDA of albendazole delivered by LF elimination programs an additional three years of community-wide MDA of albendazole will result in a lower final cross-sectional prevalence of all STH species combined as compared to a strategy of pre-SAC and SAC targeted MDA.

Objective 2.2

To compare the proportion of clusters in each of the randomization arms that interrupt transmission, defined as reaching a cluster prevalence of ≤2% for all STH species (*A. lumbricoides*, *A. duodenale*, *N. americanus* and *T. trichiura)* by quantitative PCR measured 24 months after stopping MDA.

Hypothesis: In settings where the prevalence and intensity of STH infections have been reduced through annual MDA of albendazole delivered by LF elimination programs, an additional three years of twice-yearly community-wide MDA with albendazole will result in a greater proportion of clusters achieving elimination of each STH species prevalent at baseline as compared to a strategy of pre-SAC and SAC targeted MDA measure by qPCR 24 months after stopping MDA with albendazole.

3. STUDY METHODS

3.1 TRIAL DESIGN

DeWorm3 is a multi-site community cluster randomized trial conducted in focal geographic areas in Benin, India and Malawi. At each site, 40 clusters are randomized to receive either three years of twice-annual community-wide mass drug administration (MDA) with albendazole, targeting all age groups, or the current standard of care targeted MDA strategy at the site (Table 1).

3.2 RANDOMIZATION

Within each DeWorm3 study area, clusters will be randomly assigned in a 1:1 allocation ratio to either community-wide or standard of care MDA (Table 1). Restricted randomization will be performed to ensure balance of baseline factors by arm within each country, and stratified by region in India (Jawadhu Hills, n=8 vs. Thimiri, n=32).

Table 1: Standard of care MDA strategy and eligibility for MDA by arm in each of the DeWorm3 study sites

Study site	Standard of care MDA strategy	Intervention arm community-wide MDA
Benin	Annual School-based MDA Targets pre-school age and school age children (**1-14 years old**)	Twice-annual community-wide household MDA Individuals aged **≥1 year** Not currently in the first trimester of pregnancy or acutely ill
India	Twice-annual National Deworming Days utilizing school- and community-based distribution Targets pre-school age and school age children (**1-19 years old**)	Twice-annual community-wide household MDA Individuals aged **≥1 year** Not currently in the first trimester of pregnancy or acutely ill
Malawi	Annual School-based MDA and Child Health Days Targets pre-school age and school age children (**2-19 years old**)	Twice-annual community-wide household MDA Individuals aged **≥2 years** Not currently in the first trimester of pregnancy or acutely ill

3.2.1 Cluster demarcation

Clusters were defined by each DeWorm3 Site Team, in collaboration with the Central DeWorm3 Data Team. All clusters must be contiguous and have a minimum population of 1650 individuals. Whenever possible, clusters correspond to existing administrative units.

The protocol for cluster identification consisted of the following steps:

a. Review a list of the communities (villages / neighborhoods / settlements) identified in the DeWorm3 baseline census and the total population of each according to household affiliation.
    i. If an individual community has a population within the range identified for Cluster size, that community should be considered an independent cluster and boundaries drawn based on the location of each participating household.
    ii. Where communities are too small to compose an independent cluster, GPS coordinates, field reports and the ArcMap / Google Earth map of the study area should be used to identify

communities and/or non-community-affiliated households that can reasonably be combined to form a cluster with a population within the proposed range.

iii. Where communities are too large to be considered a single cluster, GPS coordinates, field reports and an ArcMap / Google Earth map of the study area should be used to identify intuitive boundaries that result in clusters with populations within the proposed range.

b. Households not listed as belonging to a specific community should be assigned to a cluster based on geographic features and/or distance from the center of the nearest clusters to the household's GPS coordinates.

i. Where a household is equidistant between the center of two clusters, geographic features and/or the location of schools attended by children in the household may be used to make a determination.

c. Once the DeWorm3 baseline census database has been coded to reflect the cluster boundaries, this database will be sent back to the Central DeWorm3 Data Team who will review the GPS coordinates of households in each cluster to verify that households affiliated with each community fall within cluster boundaries identified. Cluster boundaries will also be reviewed in relation to geographical boundaries such as rivers and transport boundaries such as main roads.

d. Where necessary, households may be reassigned by the Central DeWorm3 Data Team to a different cluster based on cluster boundaries and GPS coordinates.

3.2.2 Cluster randomization

Clusters were randomized to community-wide versus targeted MDA stratified by country. Within each country, restricted randomization was performed to ensure balance between arms in baseline population and in factors hypothesized to be strongly associated with STH transmission intensity.

The protocol for restricted randomization was as follows:

Factors considered as part of the restricted randomization are age distribution, socio-economic status (SES), WASH access, urban / rural designation of clusters, and baseline STH prevalence as measured by Kato-Katz. Operationally, these factors and balancing criteria are defined as shown in Table 2.

Restricted randomization will be performed in Stata (StataCorp, College Station, Texas). Simulation of all possible 1:1 randomization scenarios will be performed and scenarios approved or rejected based on the balancing criteria. Validity will be checked and balancing criteria may be relaxed to ensure that the number of possible scenarios is not overly constrained, and that no two clusters are always or never assigned to the same arm; or conversely may be tightened to ensure a smaller number of possible scenarios.

Table 2: Factors chosen *a priori* to be balanced by restricted randomization

Balancing factor	Operational definition	Starting balancing criteria
Baseline STH prevalence	Prevalence of any STH by Kato-Katz, weighted by age.	1% absolute difference in mean prevalence
Age distribution	Proportion of the cluster population that are eligible for treatment in the standard of care arm (ages 1-19)	1% absolute difference in mean proportion eligible
Socio-economic status	Proportion of households with dirt floors (earth, dirt, dung)	1% absolute difference
WASH access	Proportion of households reporting no access to a toilet facility	1% absolute difference
Urban / rural designation	Cluster defined as urban or peri-urban by DeWorm3 Site Team	1 cluster difference
Total population	Total number of individuals in each arm	5,000 individuals
Cluster area	In km^2.	2.0 km^2

Once a list of approved scenarios has been prepared, DeWorm3 Site Data Teams will select the final randomization scenario by a random method of their choosing, such as a second Stata program or a public

lottery. The final scenario will be shared with the Central DeWorm3 Data Team along with a report of the randomization procedure.

3.3. SAMPLE SIZE

At each site, study areas were selected to correspond to an approximate minimum population of 80,000 to 100,000 people. The study sites were divided into 40 total clusters, 20 in the intervention arm (twice-annual community-wide MDA) and 20 in the control arm (targeted MDA). Cluster sizes vary according to administrative and / or geographic barriers but had a minimum size of 1,650 residents on the baseline census. The minimum cluster size was selected to reduce the probability of repeatedly sampling the same individuals at all three time points to <10%, allowing for 10% refusal rate and the exclusion of the Longitudinal Monitoring Cohort from the sampled population. Randomization of clusters to community-wide vs. school-age targeted MDA was performed within each site as above (Section 3.2.2).

Cross-sectional prevalence surveys at baseline and 6 months post-MDA will be conducted on a random sample of 500 individuals per cluster. The primary outcome assessment of prevalence at 24 months post-MDA will sample 1,000 people per cluster in order to precisely assess prevalence on a cluster-by-cluster basis.

3.3.1 Sample size calculations for Primary Objectives

Objective 1.1 To compare the prevalence of the predominant STH species *(A. lumbricoides, A. duodenale, N. americanus* or *T. trichiura)* as measured by quantitative PCR 24 months after stopping MDA between clusters randomized to receive twice-yearly community-wide MDA versus clusters randomized to receive standard of care pre-SAC and SAC targeted MDA.

Power calculations for a range of scenarios were conducted. The following formula outlined by Hayes and Moulton [17]:

$$c = 1 + (Z_{\alpha/2} + Z_{\beta})^2 \frac{[\pi_0(1-\pi_0) + \pi_1(1-\pi_1)][1+(m-1)\rho]}{m(\pi_0 - \pi_1)^2}$$

was used to calculate the prevalence in the control clusters (π_0) that would enable detection of a difference between arms with 80% power, given an endline prevalence in the intervention clusters (π_1) of 2% and a range of assumptions of intracluster correlation coefficient ($\rho = 0.003$ to $\rho = 0.005$), number of clusters per arm ($c = 15$ to $c = 20$) and number of people sampled per cluster ($m = 500$ or $m = 1{,}000$), assuming α=0.05. The detectable alternative π_0 ranged from 3% to 3.5%.

Power to detect differences in final prevalence between arms given plausible values for π_0, ranging from 7% to 10% (comparable to pre-intervention prevalences in each of the study sites), was then estimated using simulations conducted in R (R Center for Statistical Computing, Vienna. Austria). Each simulation assumed 1,000 individuals per cluster and a binomial distribution of STH prevalence with a mean of 2% in the intervention clusters and π_0 in the targeted arm, a range of ICC and α=0.05; 10,000 repetitions were run for each scenario. Power was ≥98% for all scenarios simulated.

Objective 1.2 To determine whether the transmission of the predominant STH species *(A. lumbricoides, A. duodenale, N. americanus* or *T. trichiura)* can be interrupted using MDA with albendazole.

Transmission interruption will be assessed at the cluster level and is defined as reaching a weighted point prevalence of ≤2% of that species as measured by quantitative PCR 24 months after stopping MDA. The proportion of clusters randomized to twice-yearly community-wide MDA versus standard of care pre-SAC and SAC-targeted MDA that interrupt transmission will be compared.

Power to detect transmission interruption and differences by arm was also estimated using simulation in R. The threshold for transmission interruption was set at 2.0% prevalence, and a cluster was considered to have interrupted transmission if prevalence could be declared to be below the threshold with 95% confidence, using a one-sided binomial test. Simulations estimated power to detect a difference in the proportion of clusters in which transmission was interrupted by arm.

Hypothesis 1.2 A
Assessment of transmission interruption in each individual cluster *i*
H_0: $\pi_i \geq 2\%$
H_A: $\pi_i < 2\%$

Hypothesis 1.2 B
Between-arm comparison of the proportion of clusters in which transmission has been interrupted
H_0: $p_0=p_1$
H_A: $p_0 \neq p_1$

Simulations assumed 20 clusters per arm, 500 or 1,000 individuals per cluster (m), and a binomial distribution of STH prevalence with a mean of 7% across clusters in the targeted arm (π_0) and π_1 in the intervention arm, a range of ICC ($\rho = 0.003$ or $\rho = 0.005$), and α=0.05; 10,000 repetitions were run for each scenario.

For a fixed mean endline prevalence of 7% in the targeted arm (π_0), power to detect a difference in the proportion of clusters eliminated varies by number of individuals sampled per cluster, ICC and endline prevalence in the intervention arm (π_1), as shown in Figure 1.

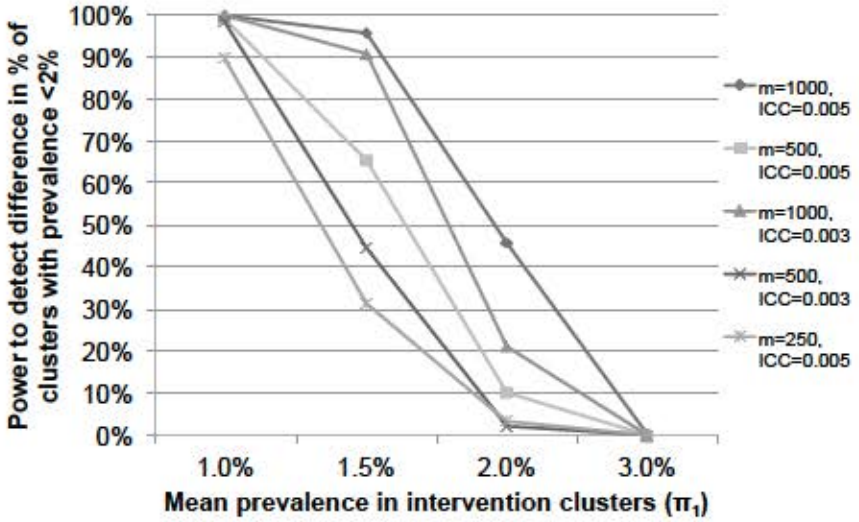

Figure 1: Power to detect a difference in the proportion of clusters achieving transmission interruption (Primary Objective 2.2.2) assuming 7% mean prevalence in the targeted clusters (π_0), by mean endline prevalence in the intervention clusters (π_1), intracluster correlation (ICC) and number of people measured per cluster (m), estimated by simulation.

These power calculations demonstrate that with 20 clusters per arm, measuring STH infection among 1,000 people per cluster at endline would provide adequate power for the primary objectives in most scenarios.

SECTION 4: STATISTICAL PRINCIPLES

4.1 CONFIDENCE INTERVALS AND P VALUES

Descriptive statistics will be presented as counts and percentages for binary, ordinal and categorical variables, and as means and standard deviations or medians and interquartile ranges for continuous variables, as appropriate based on distributions. Prevalences will use the finite population correction to account for the large proportion sampled from each cluster, and will be presented with logit Wald confidence intervals.

P-values of <0.05 will indicate statistical significance for all analyses.

4.2 ADHERENCE AND PROTOCOL DEVIATIONS

Adherence to the protocol is measured at both the cluster and individual levels. Adherence at the cluster-level refers to rounds of community-wide MDA in the intervention arm, reaching all communities in intervention clusters, being conducted according to the trial schedule laid out by each site.

Individual-level adherence refers to coverage, reach and uptake of MDA among the targeted population: school-age and preschool-age children in the standard of care arm and all age groups eligible in the intervention arm (Table 1). Coverage, reach and uptake are measured using individual treatment logs based on the most recent census in the intervention arm, and using self-report on coverage surveys in both arms (Figure 2).

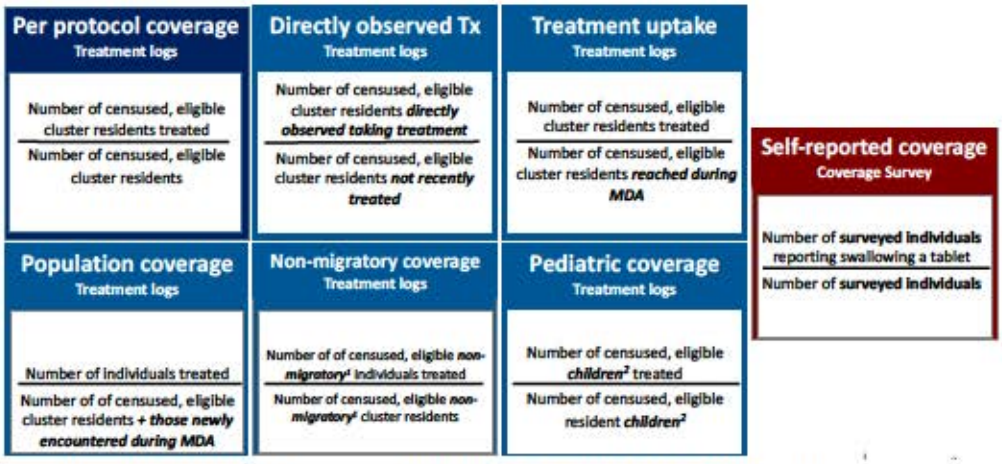

Figure 2: Treatment coverage definitions employed in DeWorm3.

Each round of MDA in the intervention arm has a target coverage and uptake of 80% of eligible study participants.

Protocol deviations, including missed or rescheduled rounds of MDA and reassignment of households or communities from one cluster to another, are documented in a Protocol Deviations Log and reviewed twice annually during DeWorm3 Data Safety and Monitoring Committee (DSMC) meetings.

4.3 ANALYSIS POPULATIONS

Analysis of DeWorm3 primary objectives will be intention-to-treat according to cluster randomization. All clusters will be considered intervention or control clusters as randomized. All residents of study clusters included on the final census will be eligible for inclusion in endline prevalence assessments and for the primary analyses will be considered exposed or unexposed based on cluster randomization, regardless of their own treatment histories.

SECTION 5: TRIAL POPULATION

5.1 ELIGIBILITY

All individuals residing in households within the study clusters will be considered part of the study population and included in the DeWorm3 baseline census and annual census updates. Cluster residents will be eligible for treatment during MDA in accordance with national guidelines at each site (Table 1). All individuals aged 1 year or older at the time of data collection (2 years or older in Malawi) will be eligible for inclusion in prevalence surveys conducted prior to the first round of MDA (Baseline), 6 months following the final round of MDA (Midline) and 24 months following the final round of MDA (Endline).

5.2 RECRUITMENT

Censuses of all individuals residing within the boundaries of each study area will be performed through door-to-door visits. Up to three attempts will be made to reach each household for inclusion in each round of the census. GPS coordinates of all dwellings, unoccupied and occupied, will be collected. The head of household, or other adult comfortable making decisions on behalf of the household, will be asked to provide consent for

inclusion in the census. Household characteristics and basic sociodemographic information on the inhabitants of each household will be collected through a survey.

The most recent census or census update will be used as a sampling frame for each round of the cross-sectional survey. Eligible participants will be randomly selected, stratified by cluster, for inclusion on a sampling list. Participants will be approached at their households and asked to participate in an in-depth survey and provide a stool sample. Adult participants will provide informed consent for their own or their children's participation, while children will provide assent, according to age guidelines in each country. Up to three attempts will be made to reach and recruit each participant on the sampling list. Data collectors will exhaust attempts to recruit individuals on the original sampling list for each cluster before being provided with supplementary lists as needed to reach the target sample size.

5.3 WITHDRAWAL

Consent for inclusion in MDA or data collection may be withdrawn during the study. Only a head of household, or the adult who originally provided consent on behalf of the household for inclusion in the census, may withdraw the entire household's consent for inclusion in the census updates and sampling lists. Individual adults may withdraw consent for their own participation in MDA or data collection activities.

5.4 BASELINE CHARACTERISTICS

Baseline characteristics collected during the DeWorm3 baseline census of all households in the study area in each site will be summarized at the site, cluster, household and individual levels (Table 4). Cluster-level variables will include those used in the restricted randomization (Table 2).

Table 4 Results of the DeWorm3 baseline census

	India		**Malawi**		**Benin**	
Study site characteristics	N (%)		N (%)		N (%)	
Geographic area of study site (km^2)						
Total number of households						
Population density <50 persons/km^2 50-249 persons/km^2 250-999 persons/km^2 ≥1000+ persons/km^2 Missing						
	N=		N=		N=	
Household characteristics	n (%) / median (IQR)		n (%) / median (IQR)		n (%) / median (IQR)	
Household size						
Owner-occupied dwelling						
Wall materials -Natural -Manmade -Other / don't know / refused						
Roofing materials -Natural -Manmade -Other / don't know / refused						
Flooring materials -Natural -Manmade -Other / don't know / refused						
WASH access						
-Sanitation -Basic facilities -Limited facilities -Unimproved facilities -No facilities (open defecation)						
-Drinking water source -Basic						

-Limited -Unimproved -Surface water						
Study population	N=		N=		N=	
Female						
Age distribution -Infants (<1 years) -Preschool-age children (1-4 years) -School age children (5-14 years) -Adults (15+ years) -Missing						
Education (school-aged children) -Attending school -Not attending school -Missing						
Highest level of education (adults) -No education -Incomplete primary school -Complete primary school -Some secondary school -Greater than secondary school -Missing / don't know / refused						
Migration -Lived outside the household the majority of the past 6 months -Slept elsewhere the night before the census						

SECTION 6: ANALYSIS

6.1 OUTCOME DEFINITIONS

The outcomes will be a) individual-level STH infection, defined as an individual's test result for the predominant STH (Objective 1.1) or any STH (Objective 2.1), and b) cluster-level transmission interruption, defined as a prevalence of the predominant STH species (Objective 1.2) or each STH species prevalent at baseline (Objective 2.2) of ≤2% with 95% confidence using a one-sided binomial test, as measured by qPCR 24 months after the final round of MDA.

We will classify each individual's infection status as negative (no amplification, Cycle threshold >40), positive, or indeterminate; primary analyses will consider indeterminate results negative, while sensitivity analyses will recategorize indeterminate results as positive. Cycle thresholds indicating a positive or indeterminate result for each species will be defined and standardized by the DeWorm3 STH Molecular Support Unit.

6.2 ANALYSIS METHODS

All analyses will be conducted separately for each site, unless otherwise specified.

6.2.1 Analysis of primary objectives

The effect of the primary exposure on endline infection with the predominant STH species at each site (Objective 1.1) will be analyzed using generalized estimating equations with Poisson family, with robust variance estimation and exchangeable correlation matrix. The primary exposure variable will be randomization arm, while the primary outcome will be individual-level STH infection status by qPCR (see Section 6.1).

Models will be adjusted for individual- and cluster-level covariates listed in Table 5. In the primary analyses, these data will be analyzed separately by site, while a secondary analysis will pool data across all sites and test for effect modification using an interaction term between study site and randomization arm.

Table 5. Individual- and cluster-level factors to be adjusted for in analysis of Objective 1.1.

Variable	Format	Data source
Cluster baseline age and gender-weighted STH prevalence	Continuous	LMC1 + CSS1
Population density within 0.5km of household	Continuous	Census 5
Household asset index quintile	Ordinal	Census 5
Water source Joint Monitoring Programme (JMP) classification	Ordinal	Census 5
Sanitation facilities JMP classification	Ordinal	Census 5
Hygiene facilities JMP classification	Ordinal	Census 5
Participant age	Continuous	Census 5
Participant gender	Categorical	Census 5
Household size	Continuous	Census 5
Migratory status	Binary	Census 5

The second primary objective of transmission interruption by cluster (Objective 1.2) will be defined as achieving an age- and gender-weighted cluster prevalence ≤2% of the predominant STH species by qPCR 24 months after the final round of MDA. The effect of the primary exposure on transmission interruption will be calculated by comparing the proportion of clusters in each arm in which transmission is successfully interrupted using a Fisher's exact test. In a sensitivity analysis, the proportion of clusters in which transmission is successfully interrupted will be compared using a weighted Chi^2 test to account for sample size contributed by cluster in case not all clusters reach enrollment targets.

6.2.2 Analysis of secondary objectives

The effect of randomization arm on endline prevalence of infection with any species of STH prevalent at baseline (2.1) will be analyzed using generalized estimating equations with Poisson family, robust variance and exchangeable correlation matrix and adjusted for individual-, household-, and cluster-level factors in Table 5 as detailed above. The effect of randomization arm on the proportion of clusters interrupting transmission of each STH species present at baseline (2.2), will be calculated by comparing the proportion of clusters in each arm in which transmission of each STH is successfully interrupted using a Fisher's exact test as detailed above.

6.2.3 Secondary analyses

Cluster-level correlates of transmission interruption (as defined by reaching a cluster prevalence of the predominant STH ≤2% 24 months after stopping MDA) will be assessed using a modified Poisson regression. Key correlates of breaking the transmission of STH will include baseline prevalence, age distribution, population migration levels, treatment coverage and adherence, rates of open defecation and population density.

Statistical software

Analyses will be conducted in R and SAS.

REFERENCES

1. Hotez, P., *Helminth Infections: Soil-transmitted Helminth Infections and Schistosomiasis*, in *Disease Control Priorities in Developing Countries*. 2006, World Bank.
2. Pullan, R.L., et al., *Global numbers of infection and disease burden of soil transmitted helminth infections in 2010*. Parasit Vectors, 2014. **7**: p. 37.
3. WHO, *Accelerating Work to Overcome the Global Impact of Neglected Tropical Diseases: A Roadmap for Implementation*. 2012, World Health Organization: Geneva
4. Organization, W.H., *Guideline: preventive chemotherapy to control soil-transmitted helminth infections in at-risk population groups* 2017, Geneva. 75.
5. Gabrielli, A.F., et al., *Preventive chemotherapy in human helminthiasis: theoretical and operational aspects.* Trans R Soc Trop Med Hyg, 2011. **105**(12): p. 683-93.
6. Jia, T.W., et al., *Soil-transmitted helminth reinfection after drug treatment: a systematic review and meta-analysis.* PLoS Negl Trop Dis, 2012. **6**(5): p. e1621.
7. Anderson, R., J. Truscott, and T.D. Hollingsworth, *The coverage and frequency of mass drug administration required to eliminate persistent transmission of soil-transmitted helminths.* Philos Trans R Soc Lond B Biol Sci, 2014. **369**(1645): p. 20130435.
8. Truscott, J.E., et al., *Can chemotherapy alone eliminate the transmission of soil transmitted helminths?* Parasit Vectors, 2014. **7**: p. 266.
9. Brooker, S.J., et al., *Global feasibility assessment of interrupting the transmission of soil-transmitted helminths: a statistical modelling study*. The Lancet Infectious Diseases, 2015. **15**(8): p. 941-950.
10. Brooker, S.J., et al., *Global feasibility assessment of interrupting the transmission of soil-transmitted helminths: a statistical modelling study*. Lancet Infect Dis, 2015. **15**(8): p. 941-50.
11. Clarke, N.E., et al., *Differential effect of mass deworming and targeted deworming for soil-transmitted helminth control in children: a systematic review and meta-analysis.* Lancet, 2017. **389**(10066): p. 287-297.