Azithromycin Reduction to Reach Elimination of Trachoma (ARRET)

Statistical Analysis Plan

1. Administrative Information

Trial registration:	NCT04185402
Funder:	National Eye Institute (UG1 EY030833)
SAP Version:	Version 2, Updated 2020-03-20
	A revision history for this document is included at the end.
Protocol Version:	Refers to Protocol / Manual of Procedures Version 1.4
Contributors:	Statisticians responsible:
	Benjamin F. Arnold, PhD [1.]
	Travis C. Porco, PhD [1]
	Principal Investigators:
	Catherine E. Oldenburg, ScD[1]
	Thomas M. Lietman, MD[1]
	Abdou Amza, MD[2]
	[1] Francis I. Proctor Foundation for Research in Ophthalmology, University of California, San Francisco, USA
	[2] Programme National de Santé Oculaire, Niamey, Niger

SAP Table of Contents

This SAP was organized following guidelines proposed in:

Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. *JAMA*. 2017;318: 2337–2343. PMID: 29260229

The companion computational notebook with underlying sample size calculations presented herein is entitled: `ARRET-sample-size-power.Rmd/.html`; it is saved in the same directory as this document.

2. Introduction

2.1. Background and rationale

[From the NIH grant application]

In this study, we will conduct a randomized evaluation of stopping versus continuing mass azithromycin distribution. Although WHO guidelines indicate that evaluation units with TF prevalence above 10% be treated for 3-5 years, this is an arbitrary threshold and trachoma may disappear in regions with declining prevalence in the absence of antibiotic intervention.

Measurement of TF is noisy, as it relies on field grades. Despite standardization, training, and certification, substantial variation between graders likely exists. Districts with measured prevalence in a given window above and below the threshold may only by chance be above or below the threshold and thus receiving treatment. If communities slightly above the 10% threshold are in reality similar to those below in true prevalence of infection, antibiotic treatment for these communities would not be expected to provide additional benefit. If trachoma is disappearing due to secular trends in improvements in sanitation and hygiene in these communities, stopping azithromycin distribution above the threshold may be possible. Early cessation of antibiotics in regions with borderline prevalence would reduce antibiotic selection pressure, which may prevent emergence of resistance in potentially pathogenic organisms.

2.2. Objectives

Specific Aim 1. Determine if local elimination of ocular *Chlamydia trachomatis* can be achieved in the absence of azithromycin intervention.

We hypothesize that stopping mass azithromycin in communities with up to 20% prevalence of TF will be non-inferior to continuation of annual mass azithromycin at 36 months.

Specific Aim 2. Assess ongoing true transmission of infection in low TF areas by evaluating alternative indicators.

We hypothesize that alternative indicators can supplement or even replace TF, and that the 95% confidence interval of the estimate for true infection from a hidden Markov model including alternative indicators of transmission will include zero (elimination).

3. Study Methods

3.1. Trial design

ARRET is a parallel, community randomized controlled trial to evaluate if mass azithromycin distribution can be discontinued in low prevalence districts. We will randomize *grappes* (approximately the size of a small village) in Niger in equal allocation (1:1) to continuation of annual mass azithromycin distribution to the entire community or stopping azithromycin treatment. For the remainder of this SAP, we will refer to *grappes* as communities. The trial will not include a placebo and will be single-masked (investigators masked). All communities will undergo identical census and monitoring. The continuation arm will receive active azithromycin

and the stopping arm will stop treatment. The trial will measure outcomes at a single timepoint, 36 months after baseline.

3.2. Randomization

Study communities will be randomized into one of the two study arms after the baseline monitoring visit. Randomization will occur after the baseline monitoring visit to reduce the potential for differential outcome assessment at the baseline visit. The randomization sequence will be generated by the trial biostatistician in San Francisco using a random number generator, without stratification or blocking. The randomization sequence will be masked from study investigators until the primary analysis is completed by the trial's biostatistician.

3.3. Sample size

The study's sample size and power analysis was based on the comparisons in Aim 1, the randomized comparisons between arms. The computational notebook that generates all ARRET sample size estimates (primary outcome) and detectable noninferiority margins (secondary outcomes) for Aim 1 is `ARRET-sample-size-power.Rmd/.html`, located in the same directory as this SAP.

3.3.1. Specific Aim 1: Noninferiority comparisons

Primary outcome: The sample size and power analysis was based on a pairwise comparison of arms. We used a standard sample size calculation for a two-sample t-test.[1] The total sample size required, *N*, is approximately:

$$N = \frac{2\left(Z_{1-\alpha/2} + Z_{1-\beta}\right)^2 S_p^2}{(\mu_1 - \mu_0 - \delta)^2}$$

where μ_1 and μ_0 are the average prevalence in treatment and stopping arms and δ is the noninferiority margin. $Z_{1-\alpha/2}$ and $Z_{1-\beta}$ are from the standard normal distribution for (two-sided) Type I error α and Type II error β. S_p is the standard deviation of the outcome (community-level prevalence). Since ARRET is a noninferiority design, we assume that $\mu_1 - \mu_0 = 0$, so the formula provides the number of communities required for a given noninferiority margin δ.

Since the ARRET trial will condition on prevalence at enrollment, we adjusted the calculation to account for correlation between the baseline measurement and the final endpoint. The adjusted standard deviation accounting for baseline measurement is $S_r = S_p\sqrt{1 - r^2}$, where *r* is the correlation between outcomes measured at baseline and follow-up.[2]

We assumed a standard deviation of 0.05 (in absolute proportion) for the community level prevalence of *C. trachomatis* infection after one year of treatment, based on previous studies.[3] We also assumed a correlation between baseline and three year values of 0.5, based on previous studies, yielding an effective standard deviation of approximately 0.043.

Under these assumptions, a sample size of *N* = 80 communities (40 per arm) will provide at least 80% power to detect our desired noninferiority margin of 3%.

Secondary Outcomes: The sample size equation above can be re-arranged to solve for the detectable noninferiority margin given a fixed sample size of *N* = 80 (with $\mu_1 - \mu_0 = 0$, as above):

$$\delta = \frac{\sqrt{2}\left(Z_{1-\alpha/2} + Z_{1-\beta}\right)S_p}{\sqrt{N}}$$

We estimated the detectable noninferiority margin for secondary outcomes TF prevalence, TI prevalence, and seroprevalence (Pgp3 and CT694 antigens). We informed the calculations for secondary outcomes using data from a cluster randomized trial in Tanzania that collected these measurements from 4,989 children ages 1-9 years from 2012 to 2015.[4,5]

TF prevalence: We assumed the between-community SD is 0.05 and the correlation between years is 0.5. The residual standard deviation is 0.043 (by coincidence, these assumptions match those for the primary outcome, *C. trachomatis* infection). The detectable noninferiority margin is approximately 3%.

TI prevalence: We assumed the between-community SD is 0.03 and the correlation between years is 0.6. The residual standard deviation is 0.024. The detectable noninferiority margin is approximately 1.7%.

Seroprevalence: We assumed the between-community SD is 0.15 and the correlation between years is 0.75. The residual standard deviation is 0.099. The detectable noninferiority margin is approximately 6.9%.

3.3.2. Specific Aim 2: Hidden Markov Models

Standard formulas for sample size estimation for hidden Markov methods are not available, but the proposed sample size can be evaluated by examination of the precision of standard sensitivity and specificity calculations (Accuracy in Parameter Estimation). Assuming 5% of the population will exhibit true positivity for a given test, we anticipate 100 true positives and 1,900 true negatives. These sample sizes provide a margin of error of approximately 4.2% for sensitivity and 0.1% for specificity (assuming true sensitivity and specificity of approximately 95%). We anticipate no loss of precision resulting from the use of the combined hidden Markov approach.

3.4. Statistical framework

The ARRET trial will use a noninferiority testing framework for the primary and secondary analyses of Aim 1. For each outcome, we will compare measures of trachoma infection (PCR, clinical, serological) between the continuation (azithromycin) arm and the stopping arm. We hypothesize that stopping treatment will be noninferior to the arm that continues treatment, with noninferiority margins specified above in the sample size section, e.g., 3% margin for the primary endpoint of *C. trachomatis* infection.

3.5. Statistical interim analyses and stopping guidance

Since the trial has only two measurement times, enrollment and final endpoint, we do not have any interim analyses planned for efficacy or futility.

3.6. Timing of the final analysis

Masked analyses will commence when all final outcomes are measured, 36 months after baseline. We will unmask the primary analysis only after the final tables and figures are complete.

3.7. Timing of outcome assessments

ARRET will measure outcomes at the baseline visit and 36 months later in repeated cross-sectional surveys.

4. Statistical Principles

4.1. Confidence intervals and *P*-values

We will construct two-sided 95% confidence intervals to test for noninferiority.

Since the trial has only two measurement times, enrollment and final endpoint, we do not have any interim analyses planned.

We will report 95% confidence intervals on our measures of effect (difference in prevalence) on each secondary outcome. We do not plan to adjust secondary outcome confidence intervals for multiplicity because we expect them to be highly correlated and together they provide corroborating evidence for or against the primary outcome hypothesis.[6]

4.2. Protocol deviations

The analysis population will include all communities randomized and all children who are age-eligible (0-9 years) at each study visit. Analyses will be intention-to-treat (ITT). We do not plan to analyze communities by different levels of adherence or treatment coverage, but we will measure it.

ARRET will measure community-level treatment coverage. Field workers will return to study community households until they distribute antibiotics to at least 80% of individuals. Section 11 of the study protocol / manual of procedures (“Study Medication”) includes additional details.

We will summarize community-level treatment coverage as the proportion of eligible children who receive treatment. The ARRET trial does not currently plan to visit study communities between baseline (month 0) and the primary endpoint (month 36). The national trachoma program will distribute azithromycin at months 12 and 24 in the continuation communities, and we will report this information in addition to study-specific treatment coverage monitoring.

5. Trial Population

5.1. Screening data

We plan to enroll a random sample of 50 children ages 0-9 years old in the enrolled study communities. We do not plan a formal assessment of representativeness of the random sample.

5.2. Eligibility

Rural communities with population sizes of 200 to 2,000 individuals will be included in the trial. Larger communities and urban and semi-urban areas will be excluded because they are substantively different from rural communities in terms of trachoma epidemiology, as they are less likely to be trachoma endemic and are not representative of rural communities where trachoma elimination is a priority.

During the mass treatment, all individuals in communities randomized to azithromycin continuation aged 1 month and up will be offered a single dose of oral azithromycin, 20mg/kg for children using height-based dosing and 1g for adults. Those under 1 month of age, pregnant, or macrolide-allergic would be offered 2 tubes of topical tetracycline ointment, to be used twice daily.

5.3. Recruitment

The trial will track and report the following populations in the CONSORT flow for the trial:

- Number of communities screened and enrolled
- Number of individuals in the community and number treated (continuation arm only)
- Number of children screened and enrolled at each measurement round
- Number of children tested for each outcome at each measurement round
- Number of children and communities analyzed for each outcome

5.4. Withdrawal/follow-up

The trial will report withdrawal of any enrolled and randomized study communities and the reason for withdrawal. Since measurement of children within communities will rely on repeated cross-sectional surveys at baseline and 36 months, children cannot be withdrawn or lost to follow-up.

5.5. Baseline patient characteristics

We will report community size (estimated total population, number of children ages 0-9 years), age and sex distributions, and baseline outcome prevalence (*C. trachomatis* infection, TF, TI, seropositivity).

6. Analysis

6.1. Outcome definitions

The primary outcome will be ocular chlamydia as assessed by PCR, measured in a population-based sample of 0-9-year-old children at 36 months after baseline.

Secondary outcomes include:

- Infectious load of chlamydia among 0-9 year-old children infected with ocular chlamydia
- Conjunctival inflammation as assessed from conjunctival photography
- Seropositivity to chlamydia based on antibody response to Pgp3 and CT694 antigens.

6.2. Analysis methods

6.2.1. Aim 1: Noninferiority comparisons

We will estimate the difference in community-level chlamydia prevalence between arms. The prespecified primary model for Specific Aim 1 is a linear model comparing community-level *C. trachomatis* prevalence among children 0-9 years, adjusting for baseline prevalence in the community, and including treatment arm as a covariate. Because group-level outcomes are used, the analysis accounts for the community randomized nature of the trial.[1]

The primary analysis will follow from the confidence interval of the estimated arm effect through the linear model, adjusting for baseline prevalence in the community. We will use a square-root transformation of infection prevalence if necessary to improve normality and heteroskedasticity.

The prespecified noninferiority margin is 3% prevalence (compared with the 5% TF threshold used in program evaluation). Our null hypothesis is that active treatment will have lower *C. trachomatis* infection of 3% or more. Evidence that prevalence in the stopping arm is not higher by more than 3% compared with the azithromycin arm will result in failure to reject a default hypothesis of inferiority (Figure 1).[7] We will estimate a two-sided 95% confidence interval.

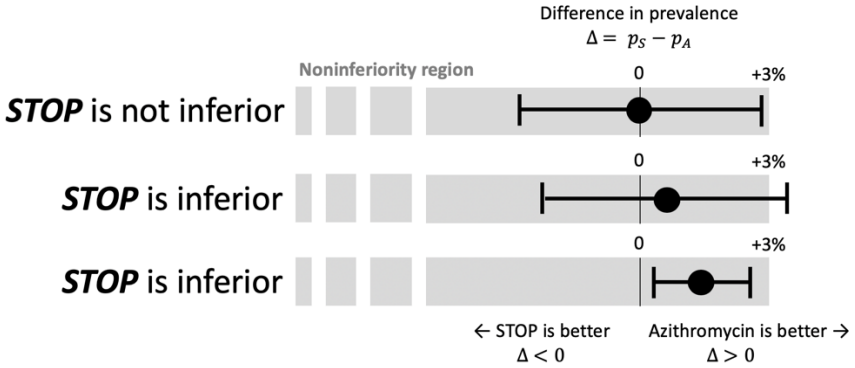

Figure 1. Hypothetical results of the trial under three scenarios: one in which stopping (*STOP*) is not inferior than azithromycin because the upper end of the two-sided 95% confidence interval of the difference between arms is less than the noninferiority margin; a second in which *STOP* is inferior because the confidence interval extends beyond the 3% noninferiority margin; and, a third in which *STOP* is inferior even though the 95% confidence interval is less than the inferiority margin because the lower interval excludes zero and azithromycin is superior, $\Delta > 0$.

Chlamydia infectious load will be modeled using clustered regression (clustering on community), using baseline community load and treatment arm as covariates. Uninfected individuals exhibit a chlamydial load of zero, so that the outcome variable is semi-continuous. We will model this with a Bernoulli-gamma mixture model (zero-inflated gamma distribution).

Inflammation and active trachoma (TF, TI) as measured by conjunctival photography are dichotomous variables at the individual level, and will yield community-level averages analogous to the primary outcome. The analysis will follow the same methods as described for the primary outcome.

Seropositivity to chlamydia will be measured from dried blood spots, including antibody responses to the antigens Pgp3 and CT694. We will dichotomize antibody responses based on

seropositivity cutoffs established through receiver-operating characteristic (ROC) curve analyses of known positive and negative specimens (conducted at the CDC WHO reference center for trachoma serology). Seropositivity at the individual level will be a dichotomous variable yielding a community-level average of seroprevalence. We will estimate seroprevalence separately by antigen and using the combined response. We will analyze two populations for serologic endpoints. At the community level, we will analyze seroprevalence among children aged 12–36 months (only those born during the duration of the trial) to assess exposure to ocular chlamydia during the study period using a model identical to the primary outcome model. We will also analyze individual-level seropositivity data in the entire age range (1–9 years for serologic measurements) to assess age-seroprevalence curves in children in continuation versus stopping communities. We will additionally estimate the seroconversion rate as measure of force of infection using age-structured seroprevalence using a generalized linear model with a complementary log-log link, age as an offset and treatment arm as a covariate.[8,9] Since this is an individual-level analysis (rather than community-level), we will estimate standard errors for the model using robust, Huber-White standard errors clustered at the community level.

The trial does not currently include prespecified subgroup analyses.

6.2.1. Aim 2 Analysis: Hidden Markov Model for Infection Prevalence

The primary analysis for Specific Aim 2 will be a hidden Markov model,[10,11] conducted at the community level using a truncated normal distribution. Analyses will be conducted in the continuation and stopping arms separately, as distribution of azithromycin likely affects the relationship between the indicators. This will allow for understanding of the best tools for surveillance both with and without azithromycin treatment. This model will use four diagnostic tests – TF, TI, ocular chlamydia, and seropositivity to CT694 and Pgp3 separately, and will not assume that any of the tests are a gold standard. The truncated normal distribution allows several options to handle the expected high density of communities with zero prevalence, including zero-inflation or a free fitted parameter. The model will incorporate delays in TF and serology, as TF follicles can take months or longer to recede and serologic measurements are long-lasting. The hidden Markov model will be fit to the prevalence of the indicators for trachoma transmission to estimate the prevalence of the true hidden prevalence of trachoma. For the test of treatment, we will assess the effect of continuation versus stopping on alternative indicators alone and in combination. Absence of an effect of treatment will suggest that there is no true infection in the community. In practice, we have found some degree of model exploration necessary, and our publications will include all explored models. Estimation will be conducted using the standard Metropolis algorithm (as we have conducted before).[12]

6.3. Missing data

Complete case analysis. We propose to analyze only the existing data. In the case of data missing from an entire community, we will simply compute the test statistic (linear model) with the available communities, conducting the permutation test with all communities. Communities with missing outcomes are still able to receive randomization allocations during the test. This would be the primary analysis reported.

Sensitivity analysis. In the event of missing outcomes for entire communities, we will assign values to the missing communities and determine how extreme the missing values would need to be for the overall conclusions to change. We will conduct this analysis by assigning the purported value, and repeating the statistical test procedure, for a range of purported values.

6.4. Additional analyses

For the primary analysis in Aim 1 that compares arms, transformations to improve heteroskedasticity and normality may be undesirable in this case, since the noninferiority margin is more interpretable on an absolute risk scale. If we have concerns about the appropriateness of the model assumptions for a community-level analysis, we will consider as an alternative a logistic regression of individual level outcomes to model the probability of infection conditional on arm and community-level baseline prevalence, with marginal standardization to estimate differences in prevalence between arms.[13] In this alternative analysis, we would estimate a 95% confidence interval of the difference using a non-parametric boostrap that resamples communities with replacement and uses the 2.5% and 97.5% percentiles of the bootstrap distribution.

6.5. Harms

Passive reporting of symptoms has been used in the past for azithromycin trials. More than 800 million doses have been distributed as part of the trachoma program, and this study will be unlikely to unveil new side effects. We intend to be vigilant, and serious adverse events suspected of being drug-related will be reported to the DSMC and to ITI within 24 hours of our notification.

6.6. Statistical software

We plan to use R statistical software (version 3.6 or later) for all analyses.

7. References

1. Hayes, R. J. & Moulton, L. H. *Cluster randomised trials*. (Taylor & Francis Group, 2017).
2. Clifton, L., Birks, J. & Clifton, D. A. Comparing different ways of calculating sample size for two independent means: A worked example. *Contemp. Clin. Trials Commun.* **13**, 100309 (2019).
3. House, J. I. *et al.* Assessment of herd protection against trachoma due to repeated mass antibiotic distributions: a cluster-randomised trial. *Lancet Lond. Engl.* **373**, 1111–1118 (2009).
4. Wilson, N. *et al.* Evaluation of a Single Dose of Azithromycin for Trachoma in Low-Prevalence Communities. *Ophthalmic Epidemiol.* **26**, 1–6 (2019).
5. Arnold, B. F. *et al.* Enteropathogen antibody dynamics and force of infection among children in low-resource settings. *bioRxiv* preprint 522920, doi: 10.1101/522920 (2019) doi:10.1101/522920.
6. Schulz, K. F. & Grimes, D. A. Multiplicity in randomised trials I: endpoints and treatments. *Lancet* **365**, 1591–1595 (2005).
7. Lesaffre, E. Superiority, equivalence, and non-inferiority trials. *Bull. NYU Hosp. Jt. Dis.* **66**, 150–154 (2008).
8. Jewell, N. P. & van der Laan, M. Generalizations of current status data with applications. *Lifetime Data Anal* **1**, 101–109 (1995).
9. Hens, N., Shkedy, Z., Aerts, M., Damme, C. F. P. V. & Beutels, P. *Modeling Infectious Disease Parameters Based on Serological and Social Contact Data*. (Springer, 2012).
10. Liu, F. *et al.* Short-term Forecasting of the Prevalence of Trachoma: Expert Opinion, Statistical Regression, versus Transmission Models. *PLoS Negl Trop Dis* **9**, e0004000 (2015).
11. Liu, F. *et al.* Short-term forecasting of the prevalence of clinical trachoma: utility of including delayed recovery and tests for infection. *Parasit Vectors* **8**, 535 (2015).
12. Lietman, T. M. *et al.* The epidemiological dynamics of infectious trachoma may facilitate elimination. *Epidemics* **3**, 119–124 (2011).
13. Muller, C. J. & MacLehose, R. F. Estimating predicted probabilities from logistic regression: different methods correspond to different target populations. *Int J Epidemiol* **43**, 962–970 (2014).

8. Revision history

Version	Date	Summary of Changes, Justification, and Timing vis-à-vis key trial events (enrollment completion, interim analyses, unmasking, etc)
1	2020-03-04	• First version adapted from the SAP developed for the NIH grant application.
2	2020-03-20	• Changed “discontinuation” arm label to “stop” label to make it more clear.