L'azithromycine pour la vie des enfants au Niger: implémentation et recherche (AVENIR)

An adaptive cluster-randomized trial to determine the optimal age group for implementation of biannual oral azithromycin distribution to reduce child mortality in Niger

Statistical Analysis Plan — Mortality and Resistance Trial

1. Administrative Information

Trial registration:	NCT04224987
Funder:	Bill & Melinda Gates Foundation (grant OPP1210548)
SAP Version:	Version 15, Updated 2023-10-17
	A revision history for this document is included at the end.
Protocol Version:	Refers to Protocol Version 25
Contributors:	Persons contributing to the SAP:
	Zijun Liu [1], Victoria Le [1], William Nguyen [1], Brittany Peterson [1]
	Statisticians responsible:
	Travis C. Porco [1], Benjamin F. Arnold [1]
	Principal Investigators:
	Thomas M. Lietman [1], Kieran O'Brien [1], Abdou Amza [2,3], Ahmed Arzika[4]

[1] Francis I. Proctor Foundation for Research in Ophthalmology, University of California, San Francisco, USA
[2] Ministère de la Santé Publique du Niger
[3] Le Programme National de Santé Oculaire (PNSO)
[4] Centre de recherche et interventions en santé publique (CRISP)

SAP Table of Contents

The content of this Statistical Analysis Plan meets the requirements stated by the US Food and Drug Administration and conforms to the American Statistical Association's Ethical Guidelines.This SAP was organized following guidelines proposed by Gamble et al. [1]

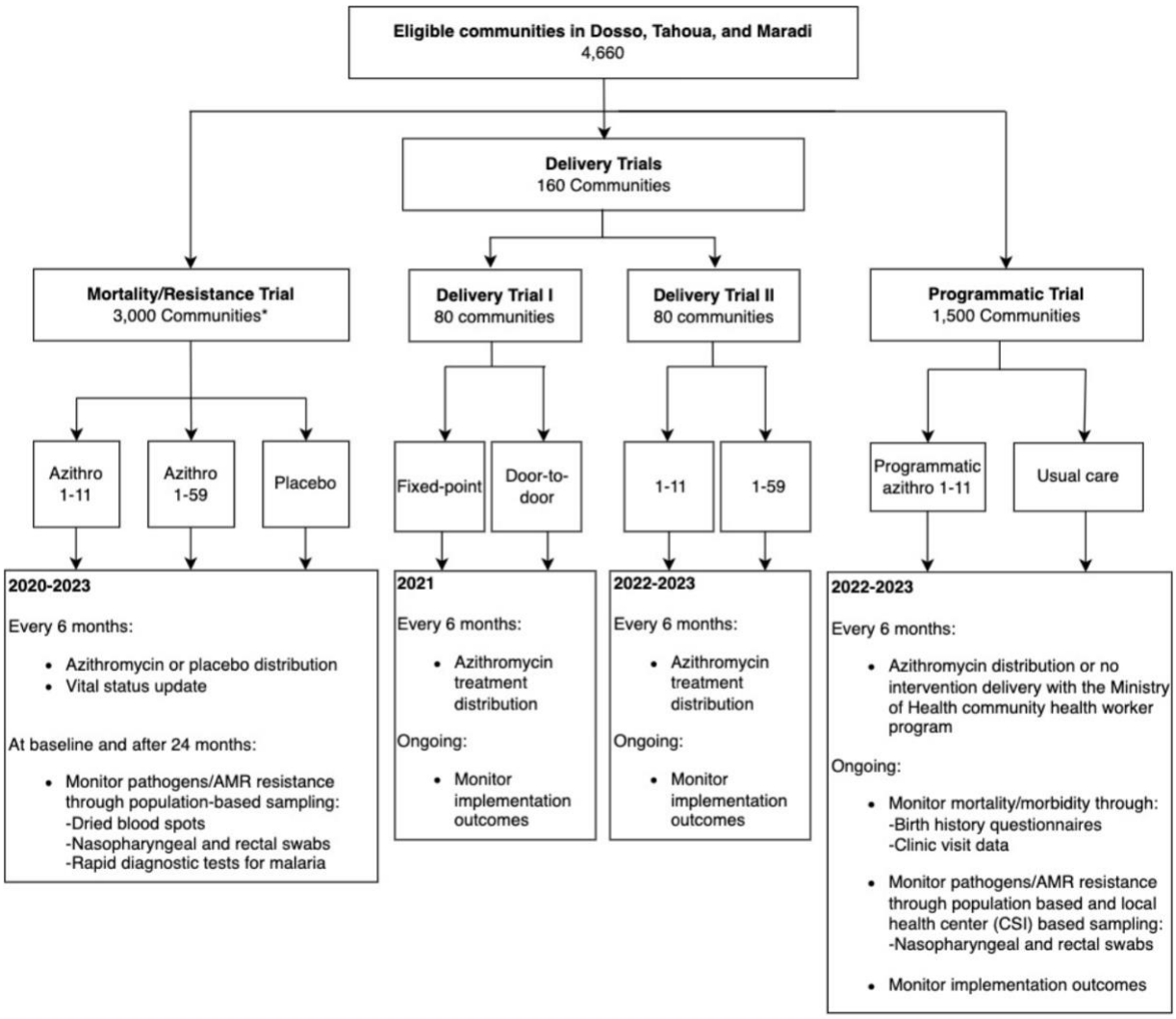
The companion computational notebook with underlying sample size calculations presented herein is entitled: AVENIR_Power_Calculation_v7.Rmd / .html; it is saved in the same directory as this document.

2. Overview

The AVENIR study includes multiple different trials, each with different objectives (Figure 1). This Statistical Analysis Plan (SAP) includes details for the Mortality and Resistance Trial. Separate SAPs cover the Delivery and Programmatic trials.

Phase I Summary of the design

Figure 1. Phase I summary of design.



*A random sample of 150 communities was selected for AMR monitoring, 60 of which contribute to the primary

Trial masking

The Mortality/Resistance trial is a double-masked, placebo controlled trial. All participants are masked to their treatement. Study investigators and field staff are masked to treatment with the

exception of the unmasked portion of the trial analysis team at UCSF. The unmasked portion of the analysis team will conduct the response adaptive allocation and pre-specified interim analyses. Unmasked members include one of the four PIs for the trial (Thomas Lietman), one of the two biostatisticians for the trial (Travis Porco), and one analyst (Zijun Liu). These members were designated in January 2022 before any adaptation or interim analyses would commence. The rationale for including this composition of the unmasked analysis team was to ensure sufficient statistical and subject matter expertise to guide the critical interim discussions and decisions for the response-adaptive trail, in coordination with the trial's Data and Safety Monitoring Committee.

The Delivery Trials and Programmatic Trial are unmasked and do not include a placebo.

3. Mortality Trial

3.1. Introduction

3.1.1. Background and rationale

This document (Statistical Analysis Plan, SAP) describes the planned analysis and reporting for the clinical trial, L'azithromycine pour la vie des enfants au Niger: implémentation et recherche (AVENIR). It includes specifications for the statistical analyses and tables to be prepared for the final Clinical Study Report. This study is a Phase IV clinical trial to compare methods to reduce childhood mortality using mass administration of azithromycin (Pfizer, CAS 83905-01-5) compared to placebo.

3.1.2. Objectives

AVENIR is designed to assess the efficacy of mass distribution of azithromycin to prevent mortality among children.

The overall objectives of this study are to:

1. Determine the optimal age group to treat with biannual oral azithromycin distribution to reduce child mortality
2. To compare selection for antimicrobial resistance across different age-based strategies for biannual azithromycin distribution
3. Demonstrate large-scale implementation of this program.

3.2. Study Methods

3.2.1. Trial design

AVENIR is designed as a large simple double-masked cluster-randomized trial with response-adaptive allocation (a platform, perpetual trial). The rationale for adaptive randomization is to allow all communities to have the opportunity to be treated, and to allow the program to evolve into the optimal treatment. If new treatment arms are added to AVENIR in the future, or if any of the current treatment arms are discontinued, the Protocol and SAP will reflect those updates.

Eligible communities in Niger will be randomly selected to participate in a mortality/resistance trial, a delivery trial or a delayed intervention.

In the mortality/resistance trial, communities will be randomized to one of three azithromycin distribution strategies: 1) **azithro 1-11:** biannual oral azithromycin to children 1-11 months old with biannual oral placebo to children 12-59 months old, 2) **azithro 1-59:** biannual oral azithromycin to children 1-59 months old, or 3) **placebo:** biannual oral placebo to children 1-59 months old. Mortality will be monitored through biannual census data collection, which will be used to adaptively allocation treatment assignments after the first year. Communities will keep their group assignment for 4 distributions. Antimicrobial resistance will be monitored in a random subsample of treated and untreated children and adults in the Dosso region. See Figure 1 for a summary of the deisgn.

Details are provided in the Protocol document, INV_002454_2019.

Mortality/Resistance Trial Interventions

Communities will be randomized to the following 3 arms, with the allocation updated after a run-in period of 1 year according to the effect of each on the mortality rate among children aged 1-59 months as determined by inter-census intervals:

- Azithromycin 1-11: biannual weight- or height-based dose of oral azithromycin suspension to children 1-11 months old and oral placebo to children 12-59 months old
- Azithromycin 1-59: biannual weight- or height-based dose of oral azithromycin suspension to children 1-59 months old
- Placebo: biannual weight- or height-based dose of oral placebo to children 1-59 months old

3.2.2. Randomization

AVENIR includes two cluster randomized trials (mortality/resistance and delivery), with grappes as the unit of randomization. Throughout we will refer to grappes as “communities”.

Mortality/Resistance Trial

At the start of the trial, communities will be allocated equally across the three arms, 1:1:1. After a run-in period of 1 year, the trial will update allocation probabilities every 6 months, corresponding to each census. Allocation probabilities between the arms will adapt based on the probability of that arm being the superior arm with respect to the mortality rate among children 1-59 months, so that newly enrolled communities will have a higher probability of being allocated to the arm with the lowest mortality rate. To minimize the impact of a possible carry-over effect of treatment in the event that a community changes arms through adaptive allocation, each community will stay with its randomized assignment for 4 distributions (approximately 2 years). Note that no community will be re-randomized by the time of the primary outcome.

Response Adaptive Allocation Algorithm

We will use the following algorithm for the adaptive allocation. Given all outcome measurements available at a given time, we will fit a negative binomial model to the data with

indicators for each arm (see section 6.2 for analysis details). From the model, we will estimate each arm's log mortality rate and its standard error. We will then draw 10,000 replicates of log mortality rates from the estimated distributions, assuming they are normally distributed, and in each draw of the joint distribution we will determine which arm has the lowest mortality rate. We will estimate the probability that each arm has the lowest mortality rate as the proportion of 10,000 replicates in which the arm has the lowest mortality rate. Following calculation of the probability that each arm has lowest mortality, we will then use a root transform of the probabilities to yield the allocation probabilities. Specifically, if p_i is the probability that arm i ($i = 1,2,3$) yields the best outcomes, then compute the allocation probabilities $a_i = p_i^{\gamma} / \sum_{j=1}^{3} p_j^{\gamma}$. We pre-specified a tempering value known only to the unmasked team to ensure masking: γ = 1/4 or 1/6 or 1/8. This transformation will help mitigate violent swings in allocation due to chance alone. Finally, we will not allow any arm's allocation probability to fall below 10% to preserve our ability to compare it against alternatives through the primary endpoint.

3.2.3. Sample size

We anticipate conducting the study in five regions of Niger: Dosso, Tahoua, Maradi, Zinder, Tillabéri; full details are given in the Protocol.

For the first two years, we expect approximately 1,400 communities in Dosso to participate for two full years (4 inter-census intervals of six month duration) and approximately 975 communities from Tahoua to participate for one year (2 inter-census intervals of six month duration) and 975 more communities from Tahoua to participate for half a year (1 inter-census interval). The sample size calculations below were informed by mortality rates and standard deviations measured in the MORDOR Niger trial [2].

Mortality rate measured as deaths per 1,000 person-years at risk will be the primary mortality outcome. The use of mortality rate protects against delays in start time across communities due to security or logistical issues, as person-time at risk naturally accounts for different follow-up times. Delayed implementation or exclusion of additional communities for security concerns could result in fewer inter-census intervals. *The sample size calculation assumes 8,525 inter-census intervals and 1,116 communities per arm.*

Mortality primary outcomes

Mortality primary outcome 1: mortality in 1-59 month olds, assessed at 2.5 years (based on at most 2 years of follow-up per community), comparing the 1-59 month and the placebo arms. Assuming a standard deviation of approximately 0.023, we find that 1,116 communities per arm will provide approximately 80% power to detect a 10% relative reduction, assuming a baseline mortality rate of 27 per thousand person-years among children 1-59 months old and an alpha of 0.05. The absolute effect size is a mortality rate difference of 2.7 deaths per 1000 person-years. This calculation is based on the standard Z-test formula for power (treating the mortality rate in each community as a continuous variable). The calculation does not involve a correction for multiple comparisons since the analysis uses a hierarchical testing approach for the three mortality outcomes, as described below.

Mortality primary outcome 2: mortality in 1-11 month olds, assessed at 2.5 years (based on at most 2 years of follow-up per community), comparing the 1-11 month arm and the placebo arm. Assuming a standard deviation of approximately 0.074, we find that 1,116 communities per arm will provide approximately 80% power to detect a 19% relative reduction, assuming a baseline

mortality rate of 45 deaths per thousand person-years among children 1-11 months old and an alpha of 0.05. The absolute effect size is a mortality rate difference of 8.8 per 1000 person-years. This calculation is based on the standard Z-test formula for power (treating the mortality rate in each community as a continuous variable). The calculation does not involve a correction for multiple comparisons since the analysis uses a hierarchical testing approach for the three mortality outcomes, as described below.

Mortality primary outcome 3: mortality in 12-59 month olds, assessed at 2.5 years (based on at most 2 years of follow-up per community), comparing the 1-11 month and 1-59 treatment arms. Assuming a standard deviation of approximately 0.022, we find that 1,116 communities per arm will provide approximately 80% power to detect a 11% relative reduction, assuming a baseline mortality rate of 24 deaths per thousand person-years among children 12-59 months old and an alpha of 0.05. This corresponds to an absolute reduction of 2.7 in the death rate per 1000 person-years. This calculation is based on the standard Z-test formula for power (treating the mortality rate in each community as a continuous variable).

Sensitivity analyses for primary outcomes based on reduced enrollment caused by security or logistical concerns: if only the first half of Tahoua communities are enrolled by 2.5 years, with 791 communities per arm, we anticipate the following detectable relative detectable effect sizes (respectively): 12%, 23% and 13%, for 1-59 months, 1-11 months and 12-59 months.

Mortality secondary outcomes

Secondary outcome 1 : mortality rate among 12-59 months comparing the 1-11 and Placebo arms. The calculation is similar to *Primary outcome 3*. So assuming a standard deviation of approximately 0.022, we find that 1,116 communities per arm will provide approximately 80% power to detect a 11% relative reduction, assuming a baseline mortality rate of 24 per thousand person-years and an alpha of 0.05. This corresponds to an absolute reduction of 2.7 in the death rate per 1000 person-years.

Secondary outcome 2: mortality rate among 1-11 months comparing the 1-11 month and 1-59 month arms. The calculation is similar to *Primary outcome 2*. So assuming a standard deviation of approximately 0.074, we find that 1,116 communities per arm will provide approximately 80% power to detect a 19% relative reduction, assuming a baseline mortality rate of 45 per thousand person-years and an alpha of 0.05. This corresponds to an absolute reduction of 8.8 in the death rate per 1000 person-years.

Sensitivity analyses for secondary outcomes based on reduced enrollment caused by security or logistical concerns: With 791 communities per arm, we anticipate the following relative detectable effect sizes (respectively): 13% and 23%, for 12-59 month olds and 1-11 month olds.

Antimicrobial Resistance (AMR) Monitoring Outcomes

AMR primary outcome 1: Prevalence of genetic determinants of resistance to macrolides including those determinants known to be found in Streptococcus pneumoniae, Streptococcus pyogenes, and Staphylococcus aureus from nasopharyngeal swabs in children 1-59 months old at 24 months from baseline. We anticipate inclusion of 50 communities per arm (150 total)

provides at least 80% power to detect an absolute difference of 8% in all age groups between any two arms, assuming:

- Baseline resistance of 30% (from MORDOR I)
- ICC of 0.045 (from multiple studies)
- 30 specimens from each of 50 communities from each arm
- Alpha of 0.05/2

AMR primary outcome 2: Load of genetic determinants of resistance to macrolides including those determinants known to be found in *Campylobacter, Salmonella* spp*., Shigella* spp., and *Escherichia coli* from rectal swabs in children 1-59 months old at 24 months from baseline. We anticipate inclusion of 50 communities per arm (150 total) provides approximately 80% power to detect a 1.4-fold difference in proportion of bacterial reads that are resistant all age groups between any two arms, assuming:

- SD of resistance load (on log scale) equal to 3.1 (MORDOR I)
- ICC of 0.02 (MORDOR I)
- 30 specimens from each of 50 communities from each arm
- Alpha of 0.05/2

AMR secondary outcome (as an example of detectable effects for other AMR secondary outcomes): For proportion of pneumococcal isolates from nasopharyngeal swabs in children 1-59 months resistant to macrolides on phenotypic testing, we anticipate inclusion of 50 communities per arm (150 total) provides approximately 80% power to detect a 8% absolute increase between any two arms, assuming:

- Baseline resistance of 10% (from MORDOR I)
- ICC of 0.045 (from multiple studies)
- 10 specimens from each of 20 communities from each arm (8 after accounting for specimen nongrowth)
- Alpha of 0.05

3.2.4. Statistical framework

The overall analysis is divided into two aspects: 1) A single frequentist analysis, conducted at the time 2.5 years after enrollment, will test the hypothesis that azithromycin MDA reduces childhood mortality. The hypothesis tests will be superiority tests (as opposed to equivalence or noninferiority analyses) and will be two-sided. 2) Ongoing analyses beyond the primary endpoint, reflecting the perpetual nature of the trial, will be conducted using a Bayesian perspective.

3.2.5. Statistical interim analyses and stopping guidance

The adaptive design will naturally change allocation, evolving toward higher allocation probabilities for arms most likely to have lowest mortality. From a frequentist perspective, we propose a interim analysis at 18 months after the first enrollment (12 months of data in each community). The analysis will be based on the Haybittle-Peto boundary which results in an interim alpha of 0.001. The alpha of the final analysis will continue to be 0.05 according to the rule.[3] The Data and Safety Monitoring Committee will be notified of the results, and if the p-value (hierarchical approach, same as primary outcome) is <0.001, then the DSMC and if appropriate, one or more PIs will meet to decide a course of action. Factors in any decision would include interest in both the 1-59 month and 1-11 month MDAs, and AMR. Early

discontinuation for other reasons (political, ethical, lack of resources, etc) would mandate final frequentist analysis at alpha of the full final prespecified value based on all available data (incorporating any alpha spending, as needed). We note that the DSMC may advise discontinuation based on evidence of harm, even in the absence of statistical significance. Bayesian inferences are unchanged by the application of interim analyses. Continuous Bayesian monitoring will be conducted.

Interim reporting

Reports will be planned for the Data and Safety Monitoring Committee. We will report the number of treatments delivered, the number of communities enrolled at each phase, and the coverage by community (i.e., process indicators). Overall mortality rates by age group (1-11, 12-59, and 1-59) will be tabulated overall in open reports.

We note that response adaptive randomization for the mortality trial will begin for communities enrolled at and after the one year mark.

3.2.6. Timing of the primary analysis

Mortality. The primary analysis will occur 2 years (rounded to the day) after the completion of enrollment. The analysis will use all the data available at that time; i.e., we do not propose to wait until phase completion for individuals enrolled and being currently followed at the 2-year mark, but rather, to use all complete data records at that moment.

AMR. Analysis of resistance outcomes will occur after the 24-month community-based samples are analyzed in the lab.

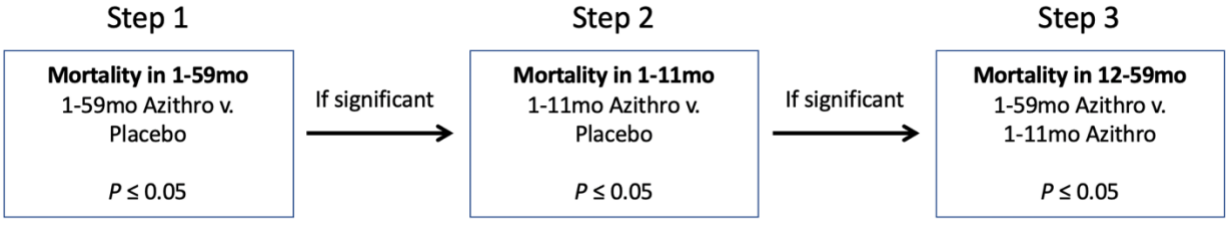
3.2.7. Timing of outcome assessments

Mortaltiy outcomes are measured at the end of each prespecified inter-census interval, intended to be six months in duration. The primary analysis will include outcomes assessed up to 2.5 years after the first enrollment. AMR outcomes are assessed at baseline and after 24 months of intervention.

3.3. Statistical Principles

3.3.1. Confidence intervals and *P*-values

The primary analysis for the mortality trial is prespecified to occur at alpha of 0.05, as follows. We propose to use a fixed-sequence, hierarchical testing approach for the three mortality primary outcomes [4]:



The staged testing approach will maximize the trial's power. Our rationale for a fixed-sequence testing approach is if there is no benefit of treating all children 1-59 mo compared with

placebo, then the trial would be unlikely to demonstrate benefit between treatment in the narrower age subgroups. Note that the hypothesis testing in the mortality trial and implementation trial will proceed independently for primary outcomes along the same sequence of tests. Regardless of the hypothesis testing results, we will report 95% confidence intervals around estimates of the incidence rate ratio for the comparisons of the trial arms.

In the mortality trial AMR substudy, we propose a second set of tests. There are two primary outcomes measured among children 1-59 months: prevalence of macrolide resistance measured in NP swabs, and load of genetic determinants of macrolide resistance measured in rectal swabs. Each primary outcome in the substudy will be examined separately at alpha of 0.05. For each outcome, we propose three pair-wise comparisons between the study groups (Placebo v 1-59mo, Placebo v 1-11mo, 1-59mo v 1-11mo). We will compute *P*-values using a permutation test with a step-down minP approach to account for multiple testing [5,6]. Additional secondary comparisons between groups in the substudy will include prevalence of genetic determinants of resistance to non-macrolide antibiotics (NP swabs) and load of genetic determinants to non-macrolide antibiotics (rectal swabs). Within each outcome, we will report *P*-values adjusted to control the false discovery rate at 5% the Benjamini-Hochberg approach [7]. For all AMR endpoints, we will additionally report unadjusted *P*-values, which will be less conservative to detect undesirable outcomes resulting from mass distribution of azithromycin.

3.3.2. Protocol deviations

The fraction of eligible censused children who accept the intervention will be reported. Communities for which this fraction falls below 60% at a specific census will be tabulated and reported.

In addition, if the fraction of children identified on a subsequent census falls below 75%, an investigation will be conducted. The number of such communities will be tabulated and reported, and sensitivity analyses omitting such communities will be reported.

Formal protocol deviations of two types are noted: (1) individual-level deviation, in which an ineligible individual is mistakenly treated, and (2) community-level deviation, in which a community is mistakenly treated with the wrong intervention. Any such events will be tabulated and reported. Note that any community which receives the wrong treatment will be analyzed according to the original randomization assignment (intention to treat).

3.3.3. Analysis Populations

Community-level data will be analyzed on an intention-to-treat basis, covering all inter-census intervals for which follow-up data are available. The random sample of communities selected for the AMR substudy will contribute to the mortality analyses.

Since AVENIR will randomize grappes before formal enrollment at baseline, the trial will analyze only grappes that are enrolled at baseline according to their assigned treatment. Grappes that are randomized but do not meet the trial's enrollment criteria will not be included in the analysis.

Modified per protocol analyses will be conducted that exclude communities with below 60% antibiotic coverage in the targeted age group. The number of such communities and their mean coverage will be reported, and such analyses will be sharply distinguished from the prespecified primary and secondary analyses.

3.4. Trial Population

3.4.1. Screening data

AVENIR will report the number of communities screened by region and the reasons communities were not enrolled based on eligibility criteria described in the next section.

3.4.2. Eligibility

This information repeats Protocol v10 section 4.3 Eligibility - Intervention.

At the community-level, eligibility includes:

Inclusion Criteria:

- Location in Dosso, Tahoua, Maradi, Zinder, or Tillabéri regions
- Population 250 to 2,499*
- Distance > 5 km from district headquarters town
- Verbal consent of community leader(s)

Exclusion criteria:

- Inaccessible or unsafe for study team
- “Quartier” designation on national census

*Population size as estimated from the most recent national census or projections

At the individual-level, eligibility includes:

Inclusion criteria:

- Age 1-59 months
- Primary residence in a study community
- Verbal consent of caregiver/guardian for study participation
- Weight ≥ 3.0 kg

Exclusion criteria:

- Known allergy to macrolides

3.4.3. Recruitment

The CONSORT flow diagram will follow the model of the MORDOR I trial [2] and will include a tabulation of the number of (a) available communities, (b) eligible communities, and (c) enrolled communities by arm.

3.4.4. Withdrawal/follow-up

Additional information will be provided. Specifically, (d) average number of individuals per community, (e) coverage and number of treatments given, and (f) number of individuals for whom end of intercensus mortality data are available.

As a large simple trial, AVENIR will not collect information regarding the specific reasons or timing for any loss to follow-up at the individual level.

3.4.5. Baseline patient characteristics

Since AVENIR is a large simple trial it will not collect extensive information about individual participants. We propose to summarize the community size (number of eligible children), child age distribution, and percentage female for enrolled communities.

3.5. Analysis

3.5.1. Outcome definitions

Mortality Trial Primary Outcomes

The trial has three primary outcomes, each a different comparison of the mortality rate (deaths per 1,000 persons at risk) at 2.5 years from the first child enrolled, with a window of 3 months.

- Mortality rate (deaths per 1,000 person-years at risk) among children 1-59 months of age compared between communities that receive azithromycin delivered to children ages 1-59 months and communities that receive placebo at 2.5 years
- Mortality rate (deaths per 1,000 person-years at risk) among children 1-11 months of age compared between communities that receive azithromycin delivered only to children ages 1-11 months and communities that receive placebo at 2.5 years
- Mortality rate (deaths per 1,000 person-years at risk) among children 12-59 months of age compared between communities that receive azithromycin delivered to children ages 1-11 months and communities that receive azithromycin delivered to children ages 1-59 months at 2.5 years

Mortality will be monitored through biannual census data collection. Trained census workers will use a custom-designed mobile application on smartphones to conduct the census. Census workers will visit each household in assigned study communities to obtain household consent for participation, and to collect demographic data on the head of household, all mothers in the household, and all children 1-59 months in the household. For eligible children, dose will be determined via weight or height and administered by the census worker. At subsequent census periods, the census workers return to each household and update the vital status of all children, indicating whether a child is alive, has died, has moved, or has an unknown status. New children and households will be added.

A child will be included as a death in the primary outcome analysis if they are present on one census and absent on the subsequent census due to death (inter-census interval). Person-time at risk will be calculated as time alive and eligible for treatment while living in the study area, with children who died contributing half of the person-time for their last inter-census interval. Children who moved, have an unknown status, or are missing from one census will contribute no person-time.

If the study team fails to measure a community for a single census due to security or logistical reasons, but re-visits the community in the following census, then person-time and outcomes for the 1-year period (2 inter-census intervals) will be counted toward the primary outcome. In this case, as with the 6-month inter-census interval, person-time at risk will be calculated as time alive and eligible for treatment for the double inter-census interval

(approximately 12 months), and children who moved, have an unknown status, or are missing from one census contributing no person-time. Children who died will contribute half of the person-time between the first census date of the double inter-census interval and their death date. Death date will be used to determine to which inter-census interval the death will belong.

Mortality Trial Secondary Outcomes

Two secondary outcomes include mortality rate comparisons at 2.5 years (±3 months) from the first child enrolled for the remaining comparisons between different age groups and arms

- Mortality rate (deaths per 1,000 person-years at risk) among children ages 12-59 months, with rates compared between communities that receive azithromycin delivered only to children ages 1-11 months and communities that receive placebo
- Mortality rate (deaths per 1,000 person-years at risk) among children ages 1-11 months, with rates compared between communities that receive azithromycin delivered only to children ages 1-11 months and communities that receive azithromycin delivered to children ages 1-59 months

Antimicrobial Resistance Substudy Primary Outcomes

The trial will have two primary outcomes for antimicrobial resistance (AMR):

- Prevalence of genetic determinants of resistance to macrolides including those determinants known to be found in *Streptococcus pneumoniae*, *Streptococcus pyogenes*, and *Staphylococcus aureus* from nasopharyngeal swabs in children 1-59 months old at 24 months from baseline
- Load of genetic determinants of resistance to macrolides including those determinants known to be found in *Campylobacter*, *Salmonella* spp., *Shigella* spp., and *Escherichia coli* from rectal swabs in children 1-59 months old after 2 years of distributions

Antimicrobial Resistance Substudy Secondary Outcomes

- Prevalence of genetic determinants of resistance to macrolides from nasopharyngeal swabs collected from children 7-12 years old after 2 years of distributions
- Prevalence of genetic determinants of resistance to macrolides from nasopharyngeal swabs collected from caregivers/guardians of eligible children after 2 years of distributions
- Prevalence of genetic determinants of resistance to other antibiotic classes in all swabs and samples listed above
- Proportion of macrolide resistant pneumococcal isolates from nasopharyngeal swabs collected from children 1-59 months
- Load of genetic determinants of resistance to all antibiotic classes from stool samples in caregivers/guardians of eligible children after 2 years of distributions

Microbiome Substudy Outcomes.

Nasopharyngeal samples will be collected from children 1-59 months of age, children 7-12 years of age, and caregivers/guardians of children 1-59 months of age as described above and in the protocol. Rectal swabs will be collected from children 1-59 months of age and stool samples will be collected from caregivers of these children as described above and in the protocol. Each age group and sample type will be used to assess the following outcomes:

- Overall microbial structure after 2 years of distribution using Manhattan distance, adjusted for baseline and analyzed with PERMANOVA to compare by arm. Euclidean distance will be used in a secondary analysis.
- Shannon's diversity index (expressed as effective number) of the microbiome after 2 years of distribution, analyzed in a linear model with a covariate for arm and baseline numbers. Simpson's index (expressed as effective number) will be used in a secondary analysis.
- Relative abundance of individual microbial taxa at the genus and species level after 2 years of distribution, adjusted for baseline and analyzed with DESeq2 to compare by arm with a Benjamini-Hochberg correction to account for the number of comparisons using a false discovery rate of 5%.

Additional Secondary Outcomes

- Mortality rate over 2.5 years by subgroups:
    - Region (Dosso, Tahoua, Maradi, Zinder, and Tillabéri)
    - Child age (1-11m, 12-24m, 24-59m)
    - Child sex
    - Community distance to the nearest Centre de Santé Intégré (CSI)
    - Number of previous community distributions
    - Timing of distribution/seasonality
    - Other interventions
        - Participation in MORDOR trials
        - Participation in and coverage of Seasonal Malaria Chemoprevention (SMC)
        - Participation in trachoma distributions of azithromycin
    - Nutritional status as defined by weight, MUAC, height interval

- Program costs as captured by routine administrative data collection during the study period and by micro-costing activities (T&M)
    - *Measured cost per dose* of azithromycin MDA in Niger for the two age groups (1 – 11 months of age and 1 – 59 months of age)
    - *Variation in measured cost per dose* of azithromycin MDA as a function of community characteristics.
    - *Modeled cost per dose* of an azithromycin MDA in new but similar settings. Values will be assigned to each *input cost category* (personnel, consumables, equipment, services, training, space), *activity* (eg, publicity, transport), and level (communities / household / child) using a fixed and variable cost perspective adjusted to the level, and adjusted for community-level characteristics.
    - Total observed azithromycin MDA program costs, by age group (1 – 11 month versus 1 – 59 month)

- Cost-effectiveness measured by estimated program costs and using measured mortality and estimated DALYs as the effectiveness outcome, compared across arms.
    - ICER for administering azithromycin to the different age groups, comparing the cost and effectiveness of 1 – 59 month vs. 1-11 month.

- Estimated 10 year national-scale effects (health and cost) of scaling up each strategy.

- Other implementation outcomes defined in the Proctor Outcomes Framework,[8][2] including feasibility and acceptability, as captured during routine administrative and monitoring data collection during the study period and by stakeholder interviews at the participant, community, and health system levels at 2.5 years (Table 3).

- Change in nutritional status over time as defined by weight, mid-upper arm circumference (MUAC), and height interval

- All-cause clinic visits among children 1-59 months of age in the study area during the study period as assessed through passive surveillance of Centre de Santé Integré records.

- Cause-specific clinic visits among children 1-59 months of age in the study area during the study period as assessed through passive surveillance of Centre de Santé Integré records, including visits with diagnoses of respiratory infection, diarrheal disease, and malaria.

- Respiratory microbiome by arm in all age groups (children 1-59 month old, children 7-12 year old and caregivers/guardians) as specified microbiome substudy primary outcomes

Other Outcomes

- Cause-specific mortality by treatment arm as defined by verbal autopsy* over 2.5 years
- Serology outcomes as assessed by dried blood spots during sample collections at baseline and 24 months
- Normalized read number of genetic determinants of resistance by antibiotic class in environmental samples at baseline and 24 months

*Depending upon resources, verbal autopsy might be conducted by UCSF or another group

MORDOR trial subgroup continuation outcomes

The AVENIR trial will include the 594 villages in Dosso region that were previously enrolled in the MORDOR I, II and III trials. This creates an opportunity to study the longer-term effects of mass azithromycin treatment on and on antimicrobial resistance.

- Longer-term carry-over effects on mortality
  - MORDOR III, beginning at month 42, re-randomized villages to receive bi-annual mass distribution of azithromycin or placebo to children 1-59 months, and followed the communities for 18 additional months (in three inter-census intervals 42-48m, 48-54m, 54-60m). From MORDOR communities allocated to the placebo arm of AVENIR, we will compare the mortality rate among children ages 1-59 months over 2.5 years in communities that received azithromycin during MORDOR III versus those that received placebo during MORDOR III. Our primary comparison of interest is among communities allocated to AVENIR's placebo arm because it will enable us to study longer-term carry-over effects of previous azithromycin treatment after discontinuation. We hypothesize

that communities that previously received azithromycin during MORDOR phase 3 will have lower mortality compared to communities that continually received placebo for the entire period. The analysis approach will follow the primary analysis, estimating incidence rate ratios using a negative binomial regression. Comparisons within additional AVENIR arms (1-59m, 1-11m) could be estimated as an exploratory analysis.

- Longer-term carry-over effects on antimicrobial resistance
  - During MORDOR III, 160 communities were tested for genetic determinants of antimicrobial resistance at month 54. This creates the opportunity to study effects of treatment history on longer-term effects for antimicrobial resistance selection. From the 160 communities in the MORDOR trial with resistance measures, the AVENIR trial will randomly select 40 communities, evenly distributed (20/20) between two treatment histories (AAAAA, AAAPP, each letter refers to a year of Azithromycin or Placebo). Within each community, we will collect a rectal swab from 30 children 1-59 months, and test the samples for genetic determinants of antimicrobial resistance using metagenomic deep sequencing, as in the resistance primary analysis. This additional substudy will enable two analyses of primary interest:
    - At enrollment in AVENIR, approximately 1 year after the end of MORDOR, we will compare the load of genetic determinants of antimicrobial resistance to macrolides (primary) and other classes of antibiotics (secondarily) between the two treatment histories. We hypothesize that load of genetic determinants will be higher among communities with the AAAAA history compared with AAAPP history. The analysis approach will follow the primary analysis for the antimicrobial resistance substudy
    - Among the 20 communities with AAAPP treatment history in MORDOR, we will compare the the load of genetic determinants of antimicrobial resistance to macrolides (primary) and other classes of antibiotics (secondarily) longitudinally over time from MORDOR 54 to AVENIR enrollment (approximately 1 year later), and again to the final measurement in AVENIR phase 1 (approximately 3 years after discontinuation of azithromycin). We hypothesize that load of genetic determinants will decline with time from last treatment.

3.5.2. Analysis methods

Mortality in the mortality/resistance trial

Our parameter of interest for comparisons of mortality between arms is the relative reduction in mortality measured by the incidence rate ratio (IRR). The primary analysis will be a Poisson regression that includes the count of deaths as the outcome and the community-level person-time at risk as an offset, with observations for each community's inter-census interval. The model will include an indicator variable for each active treatment arm. The adaptive allocation will be accounted for using pooled regression, including in the model additional indicator variables for each of three allocations in the trial (the initial, equal allocation plus 2 adaptations). For community *i* and randomization *j* (*j* = 1, 2, 3), the model will be:

$$\log(Y_{ij}) - \log(T_{ij}) = \alpha + \beta A_{ij} + \sum_{j=2}^{3} \gamma_j P_j,$$

where Y_{ij} is the number of deaths and T_{ij} is the person-time at risk in that community; A_{ij} is a treatment indicator that will vary by each comparison (details below), and P_j is an indicator for the two adaptive random allocations (the first equal allocation is the omitted category). We will

estimate 95% confidence intervals for the IRRs using bootstrap resampling (10,000 replicates) that resamples communities with replacement and re-estimates the IRRs.

For primary outcome 1 that compares mortality among 1-59 month olds, we will include in the analysis children ages 1-59 months in the 1-59 month treatment arm and placebo arm. A_{ij} will be an indicator of allocation to the 1-59 month treatment arm, and $\exp(\hat{\beta})$ will estimate the IRR.

For primary outcome 2 that compares mortality among children ages 1-11 months, we will subset the analysis to children ages 1-11 months enrolled in the 1-11 month treatment arm and placebo arm. A_{ij} will be an indicator of allocation to the 1-11 month treatment arm, and $\exp(\hat{\beta})$ will estimate the IRR.

Similarly, for primary outcome 3 that compares mortality among children ages 12-59 months, we will subset the analysis to children ages 12-59 month in the 1-11 month treatment arm and 1-59 month treatment arm. A_{ij} will be an indicator of allocation to the 1-59 month treatment arm, and $\exp(\hat{\beta})$ will estimate the IRR.

Secondary mortality analyses will be estimated similarly from the model subset to the appropriate age range.

We will estimate permutation *P*-values for each comparison using a randomization test that permutes community treatment labels within each adaptive allocation group according to the probabilities used in each allocation, holding community outcomes fixed, and will assess statistical significance using the fixed sequence, hierarchical testing procedure described above in section 3.3.

We will examine effect heterogeneity by pre-specified effect modifiers on additive and multiplicative scales [10]. We will do so by including interaction terms between effect modifiers and treatment indicators in the regression models.

Antimicrobial Resistance Substudy

The primary analysis for resistance outcomes from nasopharyngeal swabs will be ANOVA to compare the community-level proportion of individuals with any macrolide resistant genetic determinants across the three arms. The primary analysis for rectal swabs will be ANOVA to compare community-level mean load of macrolide resistant genetic determinants across the three arms. For both sets of swabs, secondary analyses examining resistance in other classes of antibiotics and other age groups will be conducted similarly. *P*-values for all resistance analyses will be computed using a permutation test, using a step down minP procedure to control for three pairwise comparisons in each primary outcome, and using a Benjamini-Hochberg correction to control the false discovery rate across additional (non-macrolide) endpoints, described above in section 4.1.

3.5.3. Missing data

Individuals with missing outcome data are assumed to contribute one half of one phase of person time to the denominator, and no counts to the numerator. Individuals are counted for each phase in which they are confirmed present at the beginning of that phase.

Communities that completely drop out during a phase will be reported, but their measurements in that phase will not contribute to the primary analysis. Earlier inter-census

intervals to which communities contributed complete data will be included in the primary analysis.

3.5.4. Additional analyses

Model Assessment

The results of additional regressions will be reported and will be clearly labeled as such and will be sharply distinguished from the primary prespecified outcome. These analyses serve two purposes: (a) assessment of how sensitive the final analytic result is to the prespecified analytic choices, and (b) facilitation of future meta-analytic or review studies. In particular, results will be reported in both absolute and relative scales. Model adequacy will be checked by examination of residuals or other goodness of fit tests as needed. Inadequate model fit will prompt us to report alternative models. As a supplemental, sensitivity analysis, we will estimate permutation *P*-values for each comparison using a randomization test proposed by Simon and Simon, which re-estimates the allocation probabilities within each permutation rather than treating them as fixed [9].

Subgroup Analyses

Pre-specified subgroup analyses for mortality constitute secondary outcomes for the trial. They include:

- Region (Dosso, Tahoua, Maradi, Zinder, Tillaberi). Rationale: examine effect heterogeneity by geographic region.
- Child age (1-11m, 12-24m, 24-59m). Rationale: examine effect heterogeneity by age within each treatment regimen
- Child sex. Rationale: consistent with US NIH reporting of sex as a biological variable.
- Community distance to the nearest Centre de Santé Intégré (CSI). Rationale: previous analysis of the MORDOR trial in Dosso, Niger found larger mortality reductions in communities further from CSIs.
- Number of previous community distributions. Rationale: examine whether effects wane with number of treatments.
- Timing of distribution / seasonality. Rationale: examine effect heterogeneity by season, as mortality could vary by season.
- Other Interventions
    - Participation in the MORDOR trials (Dosso region)
    - Seasonal Malaria Chemoprevention (SMC) in the most recent malaria season if it occurred within the three months prior to the current treatment period. Rationale: examine if there is effect heterogeneity depending on co-delivery with SMC. SMC coverage. Rationale: examine if there is effect heterogeneity depending on co-delivery with high coverage SMC.
    - Trachoma distributions of azithromycin: any distribution received by other programs during the AVENIR trial. Rationale: effect could be attenuated among communities that receive additional treatment. Number of distributions received before and during the AVENIR trial. Rationale: examine effect heterogeneity by the quantity of past and concurrent azithromycin treatments outside of the trial.

- Nutritional status as defined by weight, mid-upper arm circumference (MUAC), height interval. Rationale: determine if there is effect heterogeneity by child nutritional status measured by anthropometry.

Bayesian Analysis

To complement the primary Frequentist analysis, a Bayesian analysis of the trial will be conducted. Ongoing analyses beyond the primary endpoint, reflecting the perpetual nature of the trial, will be conducted using a Bayesian perspective. The details of the Bayesian analyses will be added to a future update of this SAP.

Costing and Cost-Effectiveness

Cost-effectiveness analysis will assess the added net costs, added health outcomes (e.g. deaths and DALYs averted), and the ICER (incremental cost-effectiveness ratio): cost per health outcome achieved by going from no intervention to 1 – 11 months, and then to 1 – 59 months as well as comparing the ICERs for census door to door, CHW door to door and CHW fixed point distribution strategies using cost estimates from the Delivery Trials described below. The health outcomes will be mortality (measured in the trial) and DALYs (calculated from mortality). Cost-effectiveness will be assessed from different perspectives: households, community, government health system, and society. Analyses will produce measures of incremental cost-effectiveness: net cost per dose delivered, per death averted, and net cost per DALY averted. The economic analysis will include two-time frames: the study period of 1 – 4 years and a longer time frame highlighting potential long-term effects such as morbidity, mortality and antimicrobial resistance. All results will be subject to sensitivity analyses: one-way (deterministic), two-way, and multi-way (probabilistic) to assess dependence of outcomes on uncertainties in input values.

The cost-effectiveness analysis will incorporate costs and effectiveness as described above. Determination of cost-effectiveness will be according to a Willingness to Pay criterion, e.g. $150/DALY averted. Costs and DALYs will be discounted at 3% per year. The study will have two time frames: the observed study period (1 – 4 years) and a longer time frame highlighting expected downstream effects and out-of-study effects (see Protocol for further discussion).

Strategies that are more costly but less or equally effective as the next alternative will be considered “dominated”, and removed from consideration, as per best practices. Strategies will be compared by cost from “no intervention” as the base case, going forward by increasing cost. Projected potential benefits in the second time frame may include delayed benefits in the intermediate term based on an observed decay curve, and potentially long-term benefits of the improved health observed during the trial (e.g., nutrition, disease episodes).

To ensure that the model and comparisons are robust, comprehensive sensitivity analyses will be conducted, including one-way, two-way, scenario, threshold and multivariate/stochastic.

3.5.5. Harms

Passive reporting of adverse effects 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 within 24 hours of our notification.

We propose no formal statistical analysis of safety events but will report adverse events by arm.

3.5.6. Statistical software

Statistical analysis will be conducted using R (R Foundation for Statistical Computing, Vienna, Austria).

4. References

1. 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. doi:10.1001/jama.2017.18556

2. Keenan JD, Bailey RL, West SK, Arzika AM, Hart J, Weaver J, et al. Azithromycin to Reduce Childhood Mortality in Sub-Saharan Africa. N Engl J Med. 2018;378: 1583–1592. doi:10.1056/NEJMoa1715474

3. Haybittle JL. Repeated assessment of results in clinical trials of cancer treatment. Br J Radiol. 1971;44: 793–797. doi:10.1259/0007-1285-44-526-793

4. Dmitrienko A, D'Agostino RB. Multiplicity Considerations in Clinical Trials. N Engl J Med. 2018;378: 2115–2122. doi:10.1056/NEJMra1709701

5. Westfall PH, Young SS. Resampling-based multiple testing: examples and methods for P-value adjustment. New York: Wiley; 1993.

6. Ge Y, Dudoit S, Speed TP. Resampling-based multiple testing for microarray data analysis. Test. 2003;12: 1–77. doi:10.1007/BF02595811

7. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol. 1995;57: 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

8. Proctor E, Hiie Silmere, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health. 2011;38: 65–76. doi:10.1007/s10477-010-0319-7

9. Simon R, Simon NR. Using randomization tests to preserve type I error with response adaptive and covariate adaptive randomization. Stat Probab Lett. 2011;81: 767–772. doi:10.1016/j.spl.2010.12.018

10. VanderWeele TJ, Knol MJ. A Tutorial on Interaction. Epidemiol Method. 2014;3: 33–72. doi:10.1515/em-2013-0005

11. Hayes RJ, Moulton LH. Cluster randomised trials. 2nd Edition. Boca Raton, FL: Taylor & Francis Group; 2017.

12. Park JJH, Detry MA, Murthy S, Guyatt G, Mills EJ. How to Use and Interpret the Results of a Platform Trial: Users' Guide to the Medical Literature. JAMA. 2022;327: 67. doi:10.1001/jama.2021.22507

13. Goodrich B, Gabry J, Ali I, Brilleman S. rstanarm: Bayesian applied regression modeling via Stan. 2022. Available: https://mc-stan.org/rstanarm

14. Gabry J, Simpson D, Vehtari A, Betancourt M, Gelman A. Visualization in Bayesian workflow. J R Stat Soc Ser A Stat Soc. 2019;182: 389–402. doi:10.1111/rssa.12378

15. Doan T, Arzika AM, Hinterwirth A, Maliki R, Zhong L, Cummings S, et al. Macrolide Resistance in MORDOR I - A Cluster-Randomized Trial in Niger. N Engl J Med. 2019;380: 2271–2273. doi:10.1056/NEJMc1901535

5. Revision history

Version	Date	Summary of Changes, Justification, and Timing vis-à-vis key trial events (enrollment completion, interim analyses, unmasking, etc)
1	2020-03-10	• First draft
2	2020-03-30	• Revision to sample size calculations (harmonization across work to date) • Revision to adaptive allocation algorithm text • Revision to analysis details (phase-stratified estimator)
3	2020-05-01	• Harmonized the order of outcomes, and list of secondary outcomes with protocol v1. Added a complete list of pre-specified secondary analyses. • Added language to allow for communities to contribute to the primary outcome if they were missed in a single census but revisited after ~12 months from the last census.
4	2020-07-27	• Added implementation trial to the SAP • Updated sample size and detectable effect calculations based on the revised design, which now includes a mortality trial and a parallel implementation trial. • Described the resistance monitoring as nested within the mortality trial. • Changed the hypothesis testing framework from Bonferroni-Holm to a fixed sequence, hierarchical testing framework to maximize study power. • Added birth history outcome measurement for the implementation trial.
5	2020-07-31	• Added Figure 1 (design overview)
6	2020-08-05	• Added interim analysis (3.5)
7	2020-09-24	Updated the design to reflect DSMC feedback, align with the Protocol: • Increased number of communities selected for mortality trial to 3,350 • Delayed start of remainder of communities until 1.5 years after the first enrollment (aligned with interim analysis) • Simplified the comparison of delivery trial approaches to a 2-arm trial and removed mortality as an outcome
8	2020-10-08	Updated outcomes and additional analyses with relevant clarifications for costing and cost-effectiveness.
9	2021-04-06	• Added MORDOR continuation substudies to the outcomes definitions in section 6.1 • Updated the multiple testing procedure planned for AMR substudy in section 4.1 • Clarified the ITT analysis will only include enrolled grappes
10	2022-02-04	• Added sensitivity analyses to Delivery Trial treatment coverage outcome in section 6.2

11	2022-05-24	• Updated the format to include 3 different sections for the 3 different trials. • Added in the Programmatic Trial section.
12	2022-05-27	• Removed details of CSI clinic visit outcome measurement and analyses until those details can be confirmed vis-à-vis field measurement and logistics.
13	2022-10-21	• Updated wording of person-time calculations for Mortality trial
14	2023-07-24	• Included in the trial overview details of study team masking (membership, timing, rationale), section 2. • Updated Figure 1 and included footnote • Updated the language of tempering roots from 1/3 or 1/4 to 1/4 or 1/6 or 1/8, section 3.2.2. • Removed redundant paragraph for masking in the section of interim Analysis, section 3.2.5 • Clarified the timing of the primary analysis, to be exactly 2 years after completion of enrollment, section 3.2.6 • Revised Mortality trial primary analysis to align notation with the final adaptive randomization schedule, including the model specification and permutation test, section 3.5.2 • Moved Delivery Trial and Programmatic Trial analysis plans to separate SAPs to make the SAP easier to navigate.
15	2023-10-17	• Updated Figure 1 to reflect the final count of communities screened and enrolled in the programmatic trial.
16	2024-09-25	• Updated AMR sample size and power calculations. The lab is going to be processing all 150 communities for the AMR outcomes, so we can include the full 150 communities in the primary AMR analyses, as opposed to the 60 that we listed originally.