Standard Operating Procedures for the LAKANA trial
SOP Data 03: Opening the trial code
Version 1.0 (2025-01-22)

1. Purpose and overview:

This SOP describes the steps for opening the trial code in LAKANA trial and obtaining the first results on the main outcome.

2. Applicability to and responsibilities of various staff members

Staff member	Responsibility
Trial Data Manager	Prepares the data for the analysis
Assistant Data Manager	Aids with the data pre-processing and conducts quality assurance procedures with Trial Statistician and Trial Data Manager
Principal Statistician	Provides instructions and insight on the statistical analysis
Trial Statistician	Conducts quality assurance with analytical scripts and conducts the analysis with the true trial codes
LAKANA Principal Investigator (PI)	Member of LAKANA Project Steering Group
LAKANA Co-Principal Investigator (Co-PI)	Member of LAKANA Project Steering Group

3. Required materials

Item	Number	Specification
Data File without explicit group information	1	File with the trial data for the analysis
Treatment Code File	1	File with letter codes that correspond to placebo/Azithromycin treatment
Merging script	1	File of code that combines treatments (placebo/Azithromycin) to corresponding letter codes and

Item	Number	Specification
		creates the three-category intervention variable
Data file with group information	1	Data file that has the intervention code included as a variable
Analytical script	1	File of code that runs the analysis
Analysis results	1	Pre-agreed tables and figures that will be produced by the Analytical script
Data Dictionary File	1	File with variables that are included in the analysis data

4. Definitions and general instructions

4.1. Definitions

4.1.1. Trial Data Manager: a LAKANA study team member who governs the data and prepares the data set to appropriate format for analysis. In the code opening event, the Trial Data Manager runs the merging scripts and the analysis script.

4.1.2. Assistant Data Manager: a LAKANA study team member who assists with data cleaning, pre-processing, and analysis procedures. In the code opening event, the Assistant Data Manager aids and gives additional oversight on the data processing procedures.

4.1.3. Trial Statistician: a LAKANA study team member who, in collaboration with the Trial Data Manager, conducts quality assurance tests on the scripts. In the code opening event, the Trial Statistician runs the analysis script jointly with the Trial Data Manager.

4.1.4. Principal Statistician: a LAKANA study team member who advises other biostatisticians and data managers on analytic strategy, statistical programming and interpretation of the trial findings and provides troubleshooting where necessary. In the code opening event, the Principal Statistician oversees the process of data analysis and reviews the first results.

4.1.5. Data File without explicit group information: File consisting of the trial data on which the analysis on mortality will be conducted on, and that doesn't have the intervention variable included. The data file will be a matrix where each row represents eligible 1-11-month-old child, and columns represent the variables described in the Data Dictionary.

4.1.6. Treatment Code File: File consisting of the letter codes used in randomization, and to which class the letters correspond to (control/Azithromycin). At the merging, explicit expressions for each letter pair code (control, biannual,

quarterly) as well as the pseudo codes for the groups (ie. “1”, ”2”, and “3”) will be included to the Treatment Code File. Example spreadsheet with mock information is presented in Appendix 1. Treatment Code File with pseudo-codes is presented in Appendix 2.

4.1.7. Merging script: Pre-prepared programmatic file of code that combines information from the Treatment Code File into the Data File without explicit group information.

4.1.8. Data file with group information: File that has gone through the merging process, and thus has the three-category intervention variable included.

4.1.9. Analytical script: Pre-prepared programmatic file of code that runs the statistical models and creates the tables that report the incidence rates, standard errors, and statistical significance between the treatment arms.

4.1.10. Analysis results: Figures and tables presenting the analysis results.

4.1.11. Data Dictionary File: File with variables that are included in the analysis data. The variables will be:

- Village Identifier (*string*)
- Child Identifier (*string*)
- Village class (*categorical, small/big village*)
- Group intervention code (*categorical)*
- Jan-Jun intervention (*categorical*)
- Jul-Dec intervention (*categorical*)
- Person-years-at-risk (*numerical*)
- Vital status at maximum one year after the first treatment (*categorical*)

Additionally, the following variables will be prepared for the analysis of effect modification:

- Age at the time of the MDA
- Sex (*dichotomous, Male/Female*)
- Weight-for-Age Z-score (WAZ)
- Seasonality (*rainy season vs non-rainy season at the time of the MDA*)
- SMC given to the child within 3 months before an MDA
- Order of MDA in the village
- District of residence
- Distance to the nearest health facility (*in km*)
- Household asset index
- Water, Sanitation, and Hygiene (WASH) index
- National outreach strategy category (standard/advanced)

5. Step-by-step procedures

Preparation

5.1. Prior to the end of data collection, the Trial Data Manager in collaboration with Assistant Data Manager and Trial Statistician will make a repeatable procedure and practice the process leading to the final analysis:

5.1.1. The pseudonymized study data for the primary outcome and subgroup analyses will be exported to a local TAU computer which has R and STATA software installed.

5.1.2. The study data will be encrypted with Cryptomator application as recommended by Tampere University.

5.1.3. The Trial Data Manager in collaboration with Assistant Data Manager and Trial Statistician pre-processes the data:

5.1.3.1. Variables defined in section 4.1.11. will be extracted from the data collection forms into a single data frame.

5.1.3.2. Only data from eligible children, that is, children who have been 1-11 months old at any time during trial, will be kept in the analysis data frame.

5.1.3.3. The pre-processed data will be saved into two file formats: a .feather formatted file and a .dta formatted file. The .feather file format will be used due to its compression properties that allow for relatively fast read-in of big files into R environment. Rationale is the same for .dta file but for STATA.

5.1.3.3.1. Two programs are used for different purposes: STATA for modeling and R for input/output processing.

5.1.4. Trial Data Manager, Trial Statistician, and Assistant Data Manager produce programmatic scripts that combine mock treatment code information from a file that is formatted in same way as the actual Treatment Code File, and that run the analyses intended for the first trial analysis.

5.1.4.1. Trial Statistician and Assistant Data Manager will both, independently, check that the merging procedure creates a file where:

1) Each eligible participant has values on variables labelled “Jan-Jun intervention” and “Jul-Dec intervention”, respectively. The variables are generated by merging information from the Treatment Code File by using the letter code information that is stored both in the data and in the Treatment Code File. Letter codes act as a merging key.

2) Each eligible participant has an intervention group identity (Control, Biannual Azithromycin, or Quarterly Azithromycin) based on the values on variables “Jan-Jun intervention” and “Jul-Dec intervention” in following manner

   a. “Placebo” value on both variables: “Control”

b. “Placebo” value on one variable and “Azithromycin” on other: “Biannual Azithromycin”

c. “Azithromycin” value on both variables: “Quarterly Azithromycin”

3) The intervention group identity is correct: participant’s intervention group identity corresponds to that which the village received in the randomization event.

4) The recoding of the intervention groups is correct:

a. Control group: 1

b. Biannual Azithromycin Group: 2

c. Quarterly Azithromycin Group: 3

5.1.4.2. At their own discretion, Trial Statistician and Assistant Data Manager can either use and review the existing merging script or generate their own to verify that the merging procedure works correctly.

5.2. In collaboration with the principal statistician, the Trial Data Manager agrees on the contents of the Analytical Script: what will be the input data, and what should the output.

Finalizing the analysis data

5.3. Once the trial has reached the end of the data collection, the Trial Data manager will produce a pre-processed Data File without explicit group information and a corresponding data dictionary. Values in the pre-processed file will be locked for the final analysis.

5.3.1. Data lock: the file into which the analysis data will be saved will be made “read-only”.

5.3.2. For preparation of any issues (file corruption, unnoticed mistakes in data etc.), a non-locked back-up version of the data will be saved to a separate location under different name.

Breaking the code

5.4. From the request of the LAKANA Principal Investigators, a representative of RTI shares the Treatment Code file with LAKANA data manager, and Pfizer will share the Treatment Code File with the Trial Statistician.

5.5. The LAKANA Trial team will convene at a pre-agreed location to produce the results from the trial.

5.6. The Trial Data Manager and the Trial Statistician will check that the two files from RTI and Pfizer match.

5.7. With approval from the PSG, the Trial Data Manager will run the Merging Script to create the intervention variable, thus creating the Data File with group information. The group intervention variable will be a categorical variable with values 1, 2, or 3.

5.8. The Trial Data Manager will review the data before the analysis by going through the pre-agreed checklist items:

1) Every participant in the analyzable data set has a final vital status

2) Every participant in the analyzable data set has PYR that is more than zero

3) Every participant in the analyzable data set has a village identity and consequently belongs to either small or big village size category

4) Every participant in the analyzable data set has intervention group identity

5) There are three intervention groups in the analyzable data set, and they are approximately in 2:4:3 ratio (Quarterly AZI: Biannual AZI: Control).

5.9. The Trial Data Manager will report if the items in the checklist will not cause any further inspections on the data or scripts.

5.10. With the permission from the PSG, The Trial Data Manager will run the Analytical Script, which has been prepared with the Trial Statistician, that produces the result output (tables and figures).

5.11. In parallel, the Trial Statistician will run their own analytical script to replicate the results to minimize the chance of human error in the Analytical Script.

6. Occupational Safety Issues

None

7. Quality Assurance / Quality Control

The scripts will be tested beforehand using mock codes by the trial team. Trial Data Manager will propose drafts of code to the Trial Statistician, and the Trial Statistician will test the scripts with the mock code file and suggests corrections to the proposition should they be necessary. The Trial Statistician and Assistant Data Manager will both, independently, check the correctness of merging data and the letter codes prior to the end of data collection.

8. Appendices and other related documents

Document number (Version)	Document content
1.0	Original document

9. Version history, authors and approvals

Version (date)	Edits to the SOP text (author)
2025-01-22	Original document (Juho Luoma) Approved by LAKANA PSG.

Appendix 1: Example of mock Treatment Code File

Letter	Random intervention
A	Placebo
B	Azithromycin
C	Placebo
E	Placebo
F	Placebo
G	Azithromycin
H	Placebo
J	Placebo
K	Azithromycin
N	Placebo
P	Azithromycin
R	Azithromycin
S	Placebo
T	Azithromycin
U	Placebo
X	Azithromycin
Y	Azithromycin
Z	Placebo

Intervention type	Jan - June group	Jan - June intervention	Jul - Dec group	Jul - Dec intervention
Control	A	Placebo	F	Placebo
Control	H	Placebo	C	Placebo
Control	E	Placebo	J	Placebo
Azi-biannual	B	Azithromycin	N	Placebo
Azi-biannual	G	Azithromycin	S	Placebo
Azi-biannual	K	Azithromycin	U	Placebo
Azi-biannual	P	Azithromycin	Z	Placebo
Azi-quarterly	T	Azithromycin	R	Azithromycin
Azi-quarterly	X	Azithromycin	Y	Azithromycin

To re-randomise, write a new number to orange cell O3 and hit tab

2

Combination number	June group	Jul - Dec group
1	T	R
2	H	C
3	X	Y
4	P	Z
5	E	J
6	B	N
7	A	F
8	K	U
9	G	S