Statistical Analysis Plan

Study alias & e-track number(s): 116682 (EPI-MALARIA-005 BOD AME)

Detailed Title:	An epidemiology study to assess *Plasmodium falciparum* parasite prevalence and malaria control measures in catchment areas of two studies pre- and post RTS,S/AS introduction (EPI-MAL-002 and EPI-MAL-003) to assess, in field conditions, vaccine benefit:risk in children in sub-Saharan Africa.
SAP Version:	***Amendment 2 Final***
SAP Date:	***08APR2020***
Scope:	All data pertaining to the above study

Statistical Analysis Plan

Study alias & e-track number(s): 116682 (EPI-MALARIA-005 BOD AME)

TABLE OF CONTENTS

PAGE

Statistical Analysis Plan	
Study alias & e-track number(s): 116682 (EPI-MALARIA-005 BOD AME)	

The SAP is divided into 2 parts:

- first part detailing the analyses to be performed (current document)
- second part: annexe (called TFL) describing the flow and format of tables, figures and listings to be annexed to the SR.

LIST OF ABBREVIATIONS

ACT	Artemisinin-based combination therapy
ATP	According to protocol
CI	Confidence interval
CTRS	Clinical Trial Registry
eCRF	Electronic case report form
GEE	Generalized estimating equations
GSK	GlaxoSmithKline
ICF	Informed consent form
IPTi	Intermittent preventative treatment in infants
IRS	Indoor residual spray
JTEG	Joint technical expert group
LL	Lower limit
LSAF	Life Science Analytics Framework
NAAT	Nucleic acid amplification test
OR	Odds Ratio
P. falciparum	*Plasmodium falciparum*
P. vivax, P. malariae, P. ovale	*Plasmodium vivax, Plasmodium malariae, Plasmodium Ovale*
PII	Personally identifiable information
QT-NASBA	Quantitative nucleic acid sequence-based amplification
QT-PCR	Quantitative polymerase chain reaction
RDT	Rapid diagnostic test
RTS	Hybrid protein comprising HBs (hepatitis B surface antibody) and CSP portions
RTS,S	Particulate antigen, containing both RTS and HBs proteins
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SAR	Statistical Analysis Report
SD	Standard Deviation
SMC	Seasonal malaria chemoprevention
TFL	Tables, Figures and Listings
UL	Upper limit
WHO	World Health Organization

Statistical Analysis Plan	
Study alias & e-track number(s): 116682 (EPI-MALARIA-005 BOD AME)	

1. DOCUMENT HISTORY

Date	Version	Description
22-SEP-2015	Final	
19-DEC-2016	Amendment 1	• Cohen's Kappa coefficient added for the comparison between microscopy and NAAT (tertiary objective) • Precisions added regarding the age derivation • Update following protocol amendment 1 (from 5 surveys to 9 surveys)
08-APR-2020	***Amendment 2***	• ***Update following protocol amendment 2:*** ○ ***The study is extended from 9 surveys to 10 surveys*** ○ ***Changes in sites participating according to sites enrolled in EPI-MAL-002 and EPI-MAL-003 studies.*** ○ ***Update the definition of vaccine eligible/ineligible subgroup based on assumptions on the duration of the EPI-MAL-005 study and dates of RTS,S vaccine implementation start.*** • ***Update of computation rules to manage discrepancies detected during the first surveys and already mentioned in previous SAR.*** • ***Additional analyses to support the EPI-MAL-003 study analysis and interpretation.*** ○ ***Demographic tables and endpoints from primary objectives will be analysed by gender, JTEG age group, vaccine eligible group and RTS,S vaccine group.*** ○ ***Adaptation of the trends analysis due to the update of the sites and the extension of the study.*** ○ ***Details of the dataset that will be incorporated into analyses from EPI-MAL-002 and EPI-MAL-003 studies are removed and will be further detailed in their respective SAPs.***

2. STUDY DESIGN

2.1. Study design overview

- ***Type of design: A multi-centric, epidemiology longitudinal cross-sectional study at centres in sub-Saharan Africa that are participating in GSK's EPI-MAL-002 and EPI-MAL-003 studies.***
- ***There will be no study vaccine administered in this epidemiology study.***
- ***Study population: Subjects 6 months to <10 years of age.***
- ***Type of study: self-contained***
- ***All medications that may influence malaria parasitaemia within 14 days prior to each survey will be recorded.***
- ***Axillary body temperature of all subjects at the time of the survey will be recorded.***
- ***Biological Samples: A capillary blood sample will be obtained for evaluation of malaria infection by blood slide and NAAT. In the event of measured fever at the time of the visit (axillary temperature ≥37.5°C) or fever reported in the last 24 hours or other symptoms/signs of clinical malaria, a rapid diagnostic test (RDT) will be conducted. If the RDT is positive, treatment will be given according to National guidelines. If a subject for whom no RDT was required is identified as being parasite positive following microscopy, National guidelines should be followed for clinical management of the subject.***
- ***Microscopy and NAAT will be used to evaluate the level of asexual and sexual parasitaemia.***
- ***Serious adverse events (SAEs) associated with the study procedure (capillary blood sampling) will be collected.***
- ***Using Geographic Information System (GIS) to determine geographical variability in MTI locally: Study areas will be mapped by villages using grid referencing. Subjects will be attributed to their village, however, to avoid PII (personally identifiable information), small villages will be grouped when the number of participants is less than 10, so that it is not possible to identify one subject from one village. Therefore the study area will be divided into segments with a minimum of 10 subjects by segment regrouping villages by proximity if needed. Villages will be grouped at the time of random sampling of the study population. A map will be drawn showing the location of numbered segments, each numbered with a unique ID code and containing one or more villages. This map will NOT be submitted to GSK, but segments will be reported by their ID code and surface (km^2) and subjects will be attributed to a segment code in the study database.***

- ***Each study site will be requested annually to provide centre specific information about interventions from the malaria control program in the study area and, if facilities are available, to provide meteorological data for the study site such as rainfall and temperature. The information will be collected in the form of a questionnaire that will be recorded in a separate database to that for subject-specific data.***
- ***Data collection: Electronic Case Report Form (eCRF).***
- ***This study will involve up to 10 annual cross-sectional surveys during malaria peak transmission with possible further extension, dependent on the duration of the EPI-MAL-002 and EPI-MAL-003 studies.***

2.2. Study groups

The following group names will be used for the statistical analyses:

Group order in tables	Group label in tables	Group definition for footnote
P. falciparum infection status measured by microscopy		
1	PF_INF	Subjects infected by *P. falciparum* parasitemia*
2	NOT_INF	Subjects not infected by *P. falciparum* parasitemia
3	UNK	Unknown
P. falciparum infection status measured by NAAT		
1	PF_INF_NAAT	Subjects infected by *P. falciparum* parasitemia*
2	NOT_INF_NAAT	Subjects not infected by *P. falciparum* parasitemia
3	UNK_NAAT	Unknown
Age classes according JTEG definition [years]		
1	0.5-4Y	6 months – 4 years at ICF date
2	5-9Y	5 – 9 years at ICF date
3	UNK	Unknown
Age classes according WHO definition [years]		
1	0.5-1Y	6 months – 1 year at ICF date
2	2-9Y	2 – 9 years at ICF date
3	UNK	Unknown
Vaccine eligible age group		
1	Eligible	Vaccine eligible age group*
2	Ineligible	Vaccine ineligible age group*
RTS,S/AS01$_E$ vaccination status		
1	Vaccinated	Children vaccinated with RTS,S/AS01$_E$ malaria vaccine*
2	Unvaccinated	Children unvaccinated with RTS,S/AS01$_E$ malaria vaccine*
Parasites density group measured by microscopy		
1	Low	Low : < 2500 parasites/ µl
2	Medium	Medium : 2500 – 9999 parasites/ µl
3	High	High : 10000 – 19999 parasites/ µl
4	Vhigh	Very high : >= 20000 parasites/ µl
5	Neg	Negative
Parasites density group measured by NAAT		
1	Low_NAAT	Low : < 1000 parasites/ µl
2	Medium_NAAT	Medium : 1000 – 10000 parasites/ µl
3	High_NAAT	High : > 10000 parasites/ µl
5	Neg_NAAT	Negative

* see section 7.4

Tables will, in general, be displayed by centre according to *P. falciparum* infection status and age classes. When it will not be the case (e.g. for some demographic tables), it will be explicitly mentioned.

According to the sites finally participating in the EPI-MAL-002 and EPI-MAL-003 studies, the following study sites names will be used for the statistical analyses:

Site n°	Vaccine exposed status	Group label used in tables	Country	Districts (eCRF denomination - Demography)
Study site				
207078	Unexposed	207078_BF_Nouna	Burkina Faso	Nouna
207079	Unexposed	207079_BF_Sapone	Burkina Faso	Sapone
207083	Exposed	207083_GH_Kintampo_EXP	Ghana	Kintampo North, Kintampo South, Nkoranza North and Techiman North
207083	Unexposed	207083_GH_Kintampo_UNEXP	Ghana	Nkoranza South and Techiman South
228607	Exposed	228607_GH_Navrongo_EXP	Ghana	Kassena-Nankana West and East
228607	Unexposed	228607_GH_Navrongo_UNEXP	Ghana	Builsa North and South
235578	Exposed	235578_KE_Gem	Kenya	Gem (Siaya)
235578	Unexposed	235578_KE_Bondo	Kenya	Bondo (Siaya)
207082	Exposed	207082_KE_Kombewa	Kenya	Kombewa
235579	Unexposed	235579_KE_Nyando	Kenya	Nyando
233149	Exposed	233149_MA_Mangochi	Malawi	Exposed Malawi 1
233149	Unexposed	233149_MA_Monkey Bay	Malawi	Unexposed Malawi 1
233150	Exposed	233150_MA_St Monfort	Malawi	Exposed Malawi 2
233150	Unexposed	233150_MA_Chikwawa	Malawi	Unexposed Malawi 2
207080	Unexposed	207080_SN_Niakhar	Senegal	Niakhar
207084	Unexposed	207084_SN_Keur Soce	Senegal	Keur Soce
207081	Unexposed	207081_TZ_Korogwe	Tanzania	Korogwe

The Senegal and Tanzania sites for which participation in the study was early terminated, were included in Survey 1 and Survey 2 only.

The Burkina Faso sites for which participation in the study was early terminated, were included in Survey 1 to Survey 4 only.

3. OBJECTIVES

As per protocol.

3.1. Co-Primary objectives

In subjects aged ≥6 months and <10 years:

- To obtain longitudinal estimates of *P. falciparum* parasite prevalence in order to characterise malaria transmission intensity in a standardised way at centres conducting the EPI-MAL-002 and EPI-MAL-003 studies before and after the introduction of the malaria vaccine RTS,S/AS01$_E$ in sub-Saharan Africa.

- To obtain longitudinal estimates of the use of malaria control interventions in centres conducting the EPI-MAL-002 and EPI-MAL-003 studies before and after the introduction of the malaria vaccine RTS,S/AS01$_E$ in sub-Saharan Africa.

3.2. Secondary objectives

In subjects aged ≥6 months and <10 years:

- To estimate trends in longitudinal estimates of the parasite prevalence of *P. falciparum* by vaccine eligible or ineligible subgroups and overall.
- To obtain longitudinal estimates of the prevalence of *Plasmodium* species other than *P. falciparum*; overall and by age group.
- To estimate longitudinal trends in receipt and timing of the third dose of DTP/ HepB/Hib and the first dose of measles EPI vaccines, at around 14 weeks and 9 months of age respectively, as appropriate by country.
- To describe changes in care seeking behaviours for reported fever or malaria in the previous 14 days.
- To assess within-site geographical heterogeneity in malaria transmission intensity.
- To describe individual malaria prevention measures and risk factors for clinical malaria according to the parasite density observed.

3.3. Tertiary objectives

In subjects aged ≥6 months and <10 years:

- To compare asexual and sexual (gametocyte) parasitaemia (qualitative- and quantitative-density) in the RTS,S/AS01$_E$ vaccinated and unvaccinated subjects.
- To compare asexual and sexual gametocyte parasitaemia (qualitative- and semi-quantitative-density) when measured by microscopy or NAAT.

4. ENDPOINTS

As per protocol.

4.1. Primary endpoints

- Occurrence of *P. falciparum* parasitaemia (using microscopy).

  Criteria/definitions: infection determined using a blood smear slide and determined using microscopy.

- Occurrence of malaria control interventions.

  Criteria/definitions: malaria control measures are mosquito net usage ***(including insecticide-treated nets [ITN] and long-lasting insecticidal nets [LLIN]), indoor***, residual spraying ***(IRS),*** seasonal malaria chemoprevention (SMC), intermittent preventative treatment in infants (IPTi), and ACT therapy received within the last 14 days.

4.2. Secondary endpoints

- Demography and history characteristics

  Criteria/definitions: gender, age, medical history.

- Occurrence of *Plasmodium* species other than *P. falciparum* (using microscopy)

  Criteria/definitions: infection with *Plasmodium* species other than *P. falciparum* determined using a blood smear slide and microscopy.

- Occurrence of uptake and timing of the third dose of DTP/HepB/Hib and the first measles EPI vaccines

  Criteria/definitions: vaccination record of receipt of dose 3 of the DTP/HepB/Hib and the first dose of the measles EPI vaccines.

- Occurrence of anti-malarial therapy

  Criteria/definitions: any anti-malarial therapy received in the last 14 days.

- Occurrence of measured fever

  Criteria/definitions: any measured fever at time of visit (axillary temperature ≥37.5°C).

- Occurrence of reported fever

  Criteria/definitions: any reported fever occurring in the last 24 hours.

- Occurrence of care seeking behaviour

  Criteria/definitions: occurrence of visits to health providers following reported fever or malaria in the previous 14 days.

- Geo-referencing characteristics

  Criteria/definitions: positioning of the subject's residence will be attributed to a segment with a unique ID from the grid referencing study area map in which the subject resides, where necessary, grouping small geographically proximate villages so that each segment has at least 10 study subjects to avoid PII, and proceeding as far as geographically appropriate.

- Occurrence of individual malaria prevention measures and risk factors

  Criteria/definitions: malaria prevention measures are repellents and local herbs not specifically recommended by the national programme and risk factors are rural/urban area, construction material for the house, floor and roof, type of eaves (open/closed), use of electricity and water source (distance from and type).

4.3. Tertiary endpoints

- Occurrence of RTS,S/AS01$_E$ vaccine doses

  Criteria/definitions: vaccination record of each dose received of the RTS,S/AS01$_E$ vaccine.

- Occurrence of *P. falciparum* parasitaemia (using NAAT)

  Criteria/definitions: detection of *P. falciparum* on dried blood spot and determined using NAAT.

- Occurrence of *P. falciparum* parasitaemia (using microscopy)

  Criteria/definitions: detection of *P. falciparum* using a blood smear slide and determined using microscopy.

- Occurrence of sexual *P. falciparum* (using NAAT/QT-NASBA)

  Criteria/definitions: detection of *P. falciparum* on dried blood spot and determined using NAAT.

5. STUDY POPULATION

As per protocol.

As a reminder, the Total cohort will include all subjects enrolled in the study. All the information for these subjects will be collected in the eCRF (after receiving signed informed consent).

The According to Protocol (ATP) cohort will include all evaluable subjects (i.e. those meeting all eligibility criteria, complying with the procedures defined in the protocol, with no elimination criteria during the study) for whom at least one laboratory result of the blood sample is available.

A detailed, comprehensive list of reasons for elimination from ATP analyses will be established at the time of data cleaning and the following elimination codes will be used.

Cohort	Elimination codes	Eli Type*
Total cohort	900	MA
ATP cohort	2010 (inclusion/exclusion criteria) and 2500 (no slides reading available)	MA

*Internal GSK database code for type of elimination code

6. STATISTICAL METHODS

As per protocol. SAS version 9.2 ***or above*** will be used for statistical analysis.

Continuous variables will be described with number of non-missing observations, mean, standard deviation (SD), median, minimum and maximum and number of missing observations. Categorical variables will be described with frequency tables; absolute numbers and percentages for each level will be given. For some categorical variables 95% confidence intervals (CIs) will also be presented.

All the analyses concerning objectives will be performed on the ATP cohort.

The analysis will be done also on the Total cohort only if we have more than 5% of eliminated subjects; except for the demography where all the tables will be performed for both cohorts.

6.1. Demography, medical history and fever

The number of subjects enrolled as well as the number excluded from the ATP analysis will be presented.

The distribution of subjects by centre will be tabulated according to *P. falciparum* infection status ***measured by microscopy*** and by both JTEG and WHO age classification.

Demographic characteristics (age at Informed Consent date, gender, previous enrolment) will be summarized by *P. falciparum* infection status ***measured by microscopy*** and overall, according to vaccine eligible status (see section 7.4.2) ***and RTS,S/AS01$_E$ vaccination status (see section 7.4.3 – only for the surveys starting after the introduction of the vaccine)***, using descriptive statistics.

Fever in the last 24 hours, fever at the visit, antimalarial or any other medication taken in the past 14 days will be tabulated by centre, and overall according to *P. falciparum* infection status and parasite density ***measured by microscopy***.

6.2. Analysis of parasite prevalence

The analysis of co-primary objective regarding the *P. falciparum* parasitemia prevalence (defined in section 7.4.5) will be done for each centre separately and by age group (WHO and JTEG) and annual age, according to vaccine eligible status (see section 7.4.2), or RTS,S/AS01$_E$ vaccination status (see section 7.4.3 – only for the surveys starting after the introduction of the vaccine) ***and by gender***. The pooled prevalence estimate will be computed considering the centre as random effect (as detailed in section 7.5). The heterogeneity among centres will be tested using Cochran's Q test based upon inverse variance weights.

This analysis of prevalence will be performed on the ATP cohort.

6.3. Analysis of Malaria control interventions

The bednet history information (sleeping under a mosquito net night before visit, new net, pierced or impregnated bednet, and how many holes) will be tabulated by centre, and overall according to *P. falciparum* infection status and parasite density ***measured by microscopy***, ***gender, vaccine eligible status (see section 7.4.2), RTS,S/AS01$_E$ vaccination status (see section 7.4.3 – only for the surveys starting after the introduction of the vaccine) and JTEG age group***.

The analysis of co-primary objective regarding the use of malaria control interventions will be done by computing the proportions of subjects using malaria control interventions among the subjects for which this information is available by centre and overall according to *P. falciparum* infection status and parasite density ***measured by microscopy*** and according to ***gender***, vaccine eligible status (see section 7.4.2), ***RTS,S/AS01$_E$ vaccination status (see section 7.4.3 – only for the surveys starting after the introduction of the vaccine) and JTEG age group***.

The CIs will be computed using the exact method.

Odds Ratio (OR) (unadjusted and adjusted) will be computed (see section ***7.6***).

This analysis will be performed on the ATP cohort

6.4. Analysis of parasite prevalence trends

The analysis of trends of parasitemia prevalence ***measured by microscopy*** (defined in section 7.7) will be performed only for the final analysis (i.e. at the last cross sectional-

surveys), and in the ATP cohort. It will be done for each centre separately and according to vaccine eligible status (see section 7.4.2).

We will compute the *P. falciparum* parasitemia prevalence (defined in section 7.4.5) for each centre separately by survey year. And the pooled prevalence estimate will be computed considering the centre as random effect (as detailed in section 7.5).

Graphs will be performed to summarize the *P. falciparum* parasitemia prevalence according to vaccine eligible status (***see section 7.4.2)*** and survey year, by centre.

6.5. Slide reading results

The analysis of slide readings (availability of slide readings, Malaria RDT results, slide methodology used, filter paper collection, number of days between date of slide reading and collection date) will be performed on the ATP cohort and will be done at the subject level (as defined in section ***7.4.4***) overall and by centre ***according to P. falciparum infection status measured by microscopy***.

6.5.1. Microscopy

6.5.1.1. Descriptive analysis

The analyses of *P. falciparum* and gametocytes will be performed using descriptive statistics for each centre separately and overall and according vaccine eligible status (see section 7.4.2) and RTS,S/AS01$_E$ vaccination status (***see section 7.4.3 – only for the surveys starting after the introduction of the vaccine).*** The density of *P. falciparum* parasitemia and gametocytes will ***also*** be evaluated.

The analyses regarding the prevalence of *Plasmodium* species other than *P. falciparum* will be done according to *P. falciparum* infection status measured by microscopy and vaccine eligible status (***see section 7.4.2)***.

The distribution of *P. falciparum* ***and density measured by microscopy*** (with CIs) will be summarized for each centre separately and overall according to bednet use, and according to fever at the visit.

The results of *P. falciparum* ***and density measured by microscopy*** will also be computed according to the gametocytes results and the malaria RDT.

6.5.1.2. Comparison analysis

The Fisher exact test will be used to perform the comparison, by RTS,S/AS01$_E$ vaccination status, of *P. falciparum* results measured by microscopy (see section 7.8).

6.5.2. NAAT results

6.5.2.1. Descriptive analysis

The analyses of *P. falciparum* and gametocytes will be performed using descriptive statistics for each centre separately and overall and according to vaccine eligible status (see section 7.4.2) and RTS,S/AS01$_E$ vaccination status ***(see section 7.4.3 – only for the surveys starting after the introduction of the vaccine).***

The results of *P. falciparum* ***measured by NAAT*** will also be computed according to the gametocytes results ***measured by NAAT***.

6.5.2.2. Comparison analysis

The Fisher exact test will be used to perform the comparison, by RTS,S/AS01$_E$ vaccination status, of *P. falciparum* results measured by NAAT (see section 7.8).

6.5.3. Comparison between microscopy and NAAT

The comparison of the *P. falciparum* ***results and density*** and gametocytes results measured by microscopy and by NAAT will be performed using the Cohen's Kappa coefficient [Cohen, 1960].

Simple Kappa coefficient will be used to compare the 2x2 contingency tables for the qualitative results (Positive/Negative, see section 7.4.4 for the details of categories).

For semi-quantitative results (see section 7.4.4 for the details of categories), simple and weighted Kappa coefficients will be used to compare the 4x4 contingency tables. The categories High : 10000 – 19999 parasites/ µl and Very high : >= 20000 parasites/ µl for the ***microscopy*** results will be pooled for this analysis.

The extent of the concordance will be assessed using the Landis and Koch scale [Landis, 1977]:

Kappa	Interpretation
0–0.20	Slight agreement
0.21–0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
0.81–1	Almost perfect agreement

6.6. Vaccination history

The vaccination history (which vaccine and age at vaccination) about DTP, Pentavalent, Measles and RTS,S/AS01$_E$ will be summarized overall and by *P. falciparum* infection status ***measured by microscopy*** using descriptive statistics.

6.7. Change in care seeking behaviours

The malaria or fever treatment taken in the past 14 days ***and details (where, when and at what time was the subject taken for treatment)***, and the malaria hospitalization will be displayed by centre, and overall according to *P. falciparum* infection status ***measured by microscopy***.

6.8. Geographical heterogeneity

The distribution of subjects by segment will be tabulated according to *P. falciparum* infection status ***measured by microscopy***.

Characteristics of segment will be summarized overall using descriptive statistics.

The analysis the *P. falciparum* parasitemia prevalence ***measured by microscopy*** (defined in section 7.4.5) will be done for each segment separately according to the centre. The heterogeneity among segment by centre will be tested using Cochran's Q test based upon inverse variance weights.

6.9. Malaria risk factors

The analysis of malaria risk factors will be performed on the ATP cohort.

The output tables for malaria risk factors (number of subjects in the same house, localisation, house information) will be done by centre and overall according to *P. falciparum* infection status and parasite density ***measured by microscopy***. OR will be computed (see section ***7.6***).

6.10. Malaria control programme at centre level

The centre specific data regarding the malaria control program will be summarized in the data listings by centre and by survey.

6.11. Analysis of SAEs related to study procedure

SAEs reported during the entire study will be tabulated per centre and overall according to *P. falciparum* infection status ***measured by microscopy*** for the Total cohort.

6.12. Output dataset

Annual fluctuations in malaria incidence occur as a result of changes in transmission intensity, which may be caused by changes in environmental factors such as rainfall or changes in usage of other malaria control interventions (use of bednets for example). Therefore, by taking into account these variations in MTI and malaria control intervention coverage, more accurate estimations of the vaccine impact on malaria disease diagnosed by RDT and/or microscopy during EPI-MAL-003 will be possible. Those estimations will be used as covariates in the models, as described in the protocol of the EPI-MAL-003.

EPI-MAL-005 is planned to run in parallel with EPI-MAL-002 and EPI-MAL-003 and will assess the following parameters: prevalence of *P. falciparum* parasitaemia, use of malaria control interventions, changes in environmental factors such as rainfall, changes in health care seeking behaviour, within-site geographical heterogeneity in MTI and individual malaria prevention measures and risk factors for malaria. Compliance with DTP-based, measles and RTS,S/AS01$_E$ vaccines will also be collected.

In order to adjust the comparison, and the estimation of the vaccine impact, an output dataset of the EPI-MAL-005 results will be built so that they could be included as covariates in the EPI-MAL-003 ***and EPI-MAL-002*** models. This dataset will consist of aggregated results by centre, and by survey. These aggregated results will ***be explicitly listed in the appendices of the SAP of the EPI-MAL-003 and Epi-MAL-002 studies with the explanation of the models used***.

7. STATISTICAL CALCULATIONS

7.1. Methodology for computing CI

All CI will be two sided 95% CI.

The exact 95% CIs for a proportion within a group will be calculated from Proc StatXact [Clopper, 1934].

The asymptotic 95% CIs for a mean within a group will be calculated.

7.2. Number of decimals

The following decimal description will be used for the analyses.

Parameters	Number of decimal digits
% of count, including LL & UL of CI	1
p-value	***4***
Minimum, maximum, range	Number of decimals in the raw data
Mean, median	Number of decimals in the raw data +1
SD	Number of decimals in the raw data +2

LL = Lower Limit UL = Upper Limit CI = Confidence Interval
SD = Standard deviation

7.3. Handling missing data

No data handling will be performed in case of missing data.

7.4. Derived and transformed data

7.4.1. Age

The age will be computed as the difference between the informed consent date and the date of birth or taken directly as reported in the CRF if the date of birth is unavailable. The age will be categorized according to the JTEG definition: 0.5-4, 5-9 years and WHO definition: 0.5-1, 2-9 years.

When the age is derived from the informed consent date and the date of birth, it could be possible that the date of birth will not be a full date. When the date will be only a year, it will be read as "*30JUNyyyy"* (except for the year of the survey – given that it is known per inclusion criteria the subject is at least 6 months old). When the date will be a month and a year, it will be read as "*15mmmyyyy"*. For these cases, the minimum age of the infant must be 6 months.

The continuous variable 'Age at IC [years]' is computed based on number of days/365.25.

7.4.1.1. Age at vaccination

The age at vaccination will be computed as the difference between the vaccination date and the date of birth.

It could be possible that the date of vaccination will not be a full date. When the date will be a month and a year, it will be read as "15mmmyyyy" except if the age becomes negative or is equal to zero, ***in which the date of birth will be used***. When the date of vaccination will be only a year, the age at vaccination will not be derived.

If 3 different dates for the DTP/HepB/Hib vaccine are entered, the last date of vaccination is selected as the date of the third dose of DTP/HepB/Hib.

7.4.2. Definition of vaccine eligible/ineligible age group

The vaccine eligible age group is defined as those subjects that on the basis of age would be eligible for RTS,S/AS01$_E$ vaccination, even if vaccine is unavailable at the time of assessment. The age will depend on the label for RTS,S/AS01$_E$ vaccination in the country. (At the time of study start, there will be no available vaccination with RTS,S/AS01$_E$).

The vaccine ineligible age group is defined as those subjects that on the basis of age would be ineligible for RTS,S/AS01$_E$ vaccination, regardless of vaccine availability at the time of assessment. The age will depend on the label for RTS,S/AS01$_E$ vaccination in the country.

The cut-off age between both subgroups will be equal to the maximum age at ***the last survey*** for a child to be able to receive a RTS,S/AS01$_E$ vaccine dose.

After introduction of the RTS,S/AS01$_E$ vaccine in 2019 and based on Country MoH eligibility guidelines, the definition of vaccine eligible/ineligible group is defined as :

- ***RTS,S/AS01$_E$ vaccine ineligible group: regroup all subjects aged more or equal to 6 years except subjects who received at least one RTS,S vaccine dose.***
- ***RTS,S/AS01$_E$ vaccine eligible group: regroup all subjects aged less than 6 years and the subjects aged more or equal to 6 years who received at least one RTS,S vaccine dose.***

7.4.3. Definition of RTS,S/ AS01$_E$ vaccination status

The vaccinated group is defined as those subjects who received at least one dose of the RTS,S/AS01$_E$ vaccine.

The unvaccinated group is defined as those subjects who received no dose of the RTS,S/AS01$_E$ vaccine.

The vaccination status will be determined only for the surveys/centres starting after the introduction of the vaccine.

Note: some vaccinated subjects were able to take part in clinical studies of the RTS,S/AS01$_E$ vaccine during phase III, so these subjects were vaccinated before the introduction of the vaccine.

7.4.4. Definition of a subject infected by a parasite – Slide reading parasitemia results at subject level

7.4.4.1. Microscopy

A subject is defined as infected by a specified parasite (*P. falciparum*, *P. malariae*, *P. vivax*, *P. ovale*) if at least two of his blood slide readings are positive for the corresponding parasite. Results of slide readings at subject level will be missing if all slide reading results are missing.

A subject is defined with presence of gametocytes when detected in at least one slide. It will be missing if all slide readings have missing information regarding the presence of Gametocytes.

The parasite density (parasites/µl) at the subject level is defined as the geometric mean of both (those positive) slide readings values if the subject's status is defined as positive. In the case of three positive slide readings, the two closest readings will be selected to calculate the geometric mean. In the case of negative parasite status, the density will be missing. If there are 3 values for one subject that have the same 'gap' between the lowest and the medium readings and between the medium and highest readings, the 2 greatest values (corresponding to the worst situation) are taken.

The parasites density will be categorized according to the following classes:

- Negative
- Low : < 2500 parasites/ µl
- Medium : 2500 – 9999 parasites/ µl
- High : 10000 – 19999 parasites/ µl
- Very high : >= 20000 parasites/ µl

For the concordances analyses the parasites density will be categorized according to the following classes:

- Negative
- Low : < 2500 parasites/ µl
- Medium : 2500 – 9999 parasites/ µl
- High and Very high : >=10000 parasites/ µl

The gametocytes density (units/ µl) at the subject level is defined as the geometric mean of positive slide readings values if the subject status is defined as positive. It will be missing otherwise.

7.4.4.2. NAAT

By the NAAT method, a subject is defined as infected by the *P. falciparum* or defined with presence of gametocytes if the NAAT reading is positive. Results of slide readings at subject level will be missing if the result is Not Done or Not Received.

The parasites density will be categorized according to the following classes:

- Negative
- Low : < 1000 parasites/ µl
- Medium : 1000 – 10000 parasites/ µl
- High : > 10000 parasites/ µl

7.4.5. Parasite prevalence

The parasite prevalence will be estimated as the proportion of subjects infected divided by the number of subjects tested. These parasites will be estimated using microscopy testing (primary and secondary analyses) and using NAAT (tertiary analyses). The CIs will be computed using the exact method.

7.4.6. Fever at the study visit

The status of fever at the study visit will be derived from the body temperature (axillary temperature). The value 'Yes' will be given if the temperature is ≥37.5°C.

When the route of t° is missing, it will be considered to be the axillary route. If the route is rectal (or tympanic with rectal conversion), then the conversion in axillary will be done, by subtracting 0.5°C (T°_Axil = T°_Rectal – 0.5).

7.4.7. Other medication

A variable 'Other medication over 14 days prior to study visit' will be derived from the eCRF if a subject was administered with a drug which was not reported as an antimalarial drug.

7.4.8. Vaccination

The vaccines DTP, Tetravalent and Pentavalent will be regrouped in one category and called 'DTP/HepB/Hib' as for the EPI-MAL-002 study.

The variable "Tradename of the vaccine" will be used to determine which vaccine is attributed to the subjects. As this variable is a free text variable, an excel file with the correct description of the vaccine was implemented by the epidemiologist.

7.5. Prevalence over all sites

The proportion of occurrence of the *P. falciparum* parasitemia over all sites will be computed based on the proportions observed in each centre. Centres will be treated as clusters and the estimates (with 95%CI) will be computed using generalized estimating equations (GEE) in order to take the correlation between subjects from each study/country into account. The GEE will be fitted using the SAS procedure GENMOD [Boos, 1992) and Rotnitzky, 1990], assuming an exchangeable correlation matrix.

7.6. Modelling of risk factors and malaria control interventions

CCI [redacted]

The potential risk factors and malaria control interventions collected will be considered.

The list of these variables can be found in the TFL. Note that some categories could be grouped at the time of the analysis in case of low representativeness.

A first table will summarise the effect of each risk factor and malaria control interventions individually on the proportion of positive episodes using unadjusted OR (with its 95% CI).

To describe the nature of the relationship between the dependent variable (probability of being *P. falciparum* positive ***measured by microscopy***) and possible explanatory variables, a multiple logistic regression model will be used. Using the Backward selection strategy (with P-value of 0.05 to stay in the model – SLSTAY = 0.05) and the method of Maximum Likelihood to estimate the parameters, the final regression equation will identify the number of possible significant risk factors. Finally, the equation will be

$$Logit(P_i) = Log\left(\frac{P_i}{1-P_i}\right) = \alpha + \sum_j \beta_j X_j$$

where

P_i = probability for a subject to be infected

X_j = significant risk factors

A second table will give the details for the estimated coefficient, the adjusted OR, and the associated p-value ***by using the variable site as cluster***.

The adjusted analysis will be done on one hand for risk factors and for malaria control interventions separately and on the other hand for all variables as a whole.

For the adjusted model, the variables depending on another variable will be adapted to avoid introducing missing values in the model :

- ***If the variable "Antimalarial or any other medication within 14 days prior the study visit" = No, then the value for the variables "Antimalarial drug" and "Other medication over 14 days prior the study visit" is assigned to No.***
- ***If the variable "Indoor residual spray used in past 12 months to spray interior walls of this part of the house" = No, then the value for the variable "Use of indoor residual spray – number of months age" = No Residual Spray.***
- ***If the variable "subject sleep under a mosquito net last night" = No, then the value for the variables "new net (less than 1 year)" "Impregnated Bednet" and "Pierced/torn Bednet" = No Mosquito Net.***
- ***If the variable "Pierced/torn Bednet" = No, then the value for the variable "Number of holes" = No Pierced Bednet.***
- ***The categories of the variable "Use of indoor residual spray – number of months ago" will be regrouped to:***
  - 1-2 months
  - 3-4 months
  - > 4 months
  - No Residual Spray.

7.7. Parasitemia prevalence trends

To determine if a parasitemia prevalence trends may exist across the surveys, a Cochran-Armitage trend test is computed.

As some sites were early terminated during the study and others started after the RTS,S/AS01$_E$ vaccine introduction, the Cochran-Armitage trend test will be computed only on centres with results for at least 3 surveys.

The Cochran-Armitage test is a method of directing chi squared tests toward narrow alternatives. The test is sensitive to the linearity between response variable and experimental variables and detects trends. [Cochran, 1954, Armitage, 1955].

To compute the OR of Survey year according to vaccine eligible status (***see section 7.4.2)*** and for each ***site***, a simple logistic regression model will be used (see section 7.6).

The survey's reference category will be the first survey of the centre.

7.8. Fisher exact test

The comparisons of proportions for categorical variables will be done with Exact Fisher test. The *p*-values less than 0.05 will be used as an indicator that a difference between the two groups compared may exist. However since *p*-values will not be adjusted for multiplicity comparisons, they will have to be interpreted with caution.

7.9. Rapid Diagnostic Test (RDT)

According to protocol, an RDT is performed when the subject has fever at the time of the visit (axillary temperature ≥37.5°C) or fever reported in the last 24 hours or, after amendment 1 of the protocol, other symptoms/signs of clinical malaria. However, several protocol deviations were detected during the first surveys concerning the RDT use, i.e., some centres performed an RDT on non-eligible subjects.
Due to this deviation, additional tables will be computed on the RDT results restricted to subjects with fever at first surveys and to subjects with fever or symptoms/signs of clinical malaria at surveys following amendment 1 of the protocol.

7.10. Malaria risk factors adaptation

An inconsistency was identified during the analysis of survey 2 between the 2 nested variables "What are the main types of windows in the house?" and "Are there nets on the main windows of the house?" (household questionnaire).

For most of the subjects living in a house with "no window", the category "Nets not present" was ticked for the nested variable describing net characteristics.

As a priori living in a house with "no window" could be considered as a protective factor when living in a house with "Nets not present" could be considered as a risk factor for malaria; the team decided to attribute to all subjects living in a house with 'No Window' the value 'Other' for the variable describing net characteristics.

8. CONDUCT OF ANALYSES

8.1. Sequence of analyses

10 analyses will be performed ***with possible further extension, depending on the duration of the EPI-MAL-002 and EPI-MAL-003 studies.***

	(*LSAF* sub-folder)	
Information collected in the first annual cross sectional survey	ANALYSIS_E1_02	Cross Survey 1
Information collected in the second annual cross sectional survey	ANALYSIS_E1_03	Cross Survey 2
Information collected in the third annual cross sectional survey	ANALYSIS_E1_04	Cross Survey 3
Information collected in the fourth annual cross sectional survey	ANALYSIS_E1_05	Cross Survey 4
Information collected in the fifth annual cross sectional survey	ANALYSIS_E1_07	Cross Survey 5
Information collected in the sixth annual cross sectional survey	ANALYSIS_E1_08	Cross Survey 6
Information collected in the seventh annual cross sectional survey	ANALYSIS_E1_09	Cross Survey 7
Information collected in the eighth annual cross sectional survey	ANALYSIS_E1_10	Cross Survey 8
Information collected in the ninth annual cross sectional survey	***ANALYSIS_E1_11***	***Cross Survey 9***
Information collected in the ***tenth*** annual cross sectional survey and trend analysis over the ***10*** years.	ANALYSIS_E1_01	Cross Survey ***10*** and Trend analysis
Dataset on EPI-MAL-005 results used as covariates in EPI-MAL-002 analyses	***ANALYSIS_E1_12***	***NA***
Dataset on EPI-MAL-005 results used as covariates in EPI-MAL-003 analyses	***ANALYSIS_E1_13***	***NA***

The analysis ID associated in Life Science Analytics Framework (LSAF) could be updated according the needs of the study.

The tables or graphs that will be performed only in the last survey are clearly identified in the TFL.

8.2. Statistical considerations for interim analyses

Not applicable

9. CHANGES FROM PLANNED ANALYSES

CCI

[redacted]

10. REFERENCES

Armitage, P (1955). "Tests for Linear Trends in Proportions and Frequencies". Biometrics (International Biometric Society) 11 (3): 375–386.

Boos, D. On Generalized Score Tests. *The American Statistician* 1992; 46, 327–333.

Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of binomial. *Biometrika* 1934; 26: 404-413.

Cochran, WG (1954). "Some methods for strengthening the common chi-squared tests". Biometrics (International Biometric Society) 10 (4): 417–451.

Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;(20):27-46.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics.mars 1977;33(1):159-74.

Rotnitzky, A. and Jewell, N. P. Hypothesis Testing of Regression Parameters in Semiparametric Generalized Linear Models for Cluster Correlated Data. *Biometrika* 1990; 77, 485–497.