Data analysis plan

Subpatent Malaria in Pregnancy Study Group

Version 1.0

May 2019

WorldWide Antimalarial Resistance Network (WWARN)

WWARN Logo (new version)

**Suggested citation:** Study protocol and data analysis plan, Subpatent in Malaria Study Group: exploring the relationship between subpatent *Plasmodium falciparum* malaria infection and pregnancy outcome.

Version History

|  |  |  |
| --- | --- | --- |
| Version number | Revision(s) & reason for amendment | Release date |
| v1.0 |  | 13.05.2019 |
|  |  |  |
|  |  |  |

WorldWide Antimalarial Resistance Network (WWARN)

[www.wwarn.org](http://www.wwarn.org)

Contents

[Introduction 5](#_Toc8636695)

[Aims and Objectives of the Study 5](#_Toc8636696)

[Project 1. Evaluation of the association between subpatent malaria and adverse pregnancy outcomes 6](#_Toc8636697)

[1. Eligibility criteria for inclusion in pooled analysis 6](#_Toc8636698)

[1.1. Essential Data 6](#_Toc8636699)

[1.2. Desirable Data (not required for inclusion) 7](#_Toc8636700)

[1.3. Study Exclusion criteria 7](#_Toc8636701)

[1.4. Participant exclusion 7](#_Toc8636702)

[2. Methodologies 7](#_Toc8636703)

[2.1. Data standardisation 7](#_Toc8636704)

[2.2. Additional data 8](#_Toc8636705)

[3. Exposures 8](#_Toc8636706)

[4. Outcomes 9](#_Toc8636707)

[5. Analysis Population 11](#_Toc8636708)

[6. Covariates Examined 11](#_Toc8636709)

[7. Outline of Statistical Analysis 12](#_Toc8636710)

[7.1. Flow chart 12](#_Toc8636711)

[7.2. Baseline characteristics of the participants included in the IPD analysis 12](#_Toc8636712)

[7.3. Meta-Analysis 12](#_Toc8636713)

[8. Statistical Methodology 16](#_Toc8636714)

[8.1. Flow chart 16](#_Toc8636715)

[8.2. Descriptive statistics 16](#_Toc8636716)

[8.3. Analytical Methods 16](#_Toc8636717)

[8.4. Assessment of heterogeneity between sites 16](#_Toc8636718)

[8.5. Model selection for determinants 16](#_Toc8636719)

[8.6. Missing data 17](#_Toc8636720)

[8.7. Sensitivity Analysis 17](#_Toc8636721)

[8.8. Risk of bias 18](#_Toc8636722)

[9. Tools 18](#_Toc8636723)

[10. References 18](#_Toc8636724)

Abbreviations in alphabetical order

|  |  |
| --- | --- |
| Dhps | Dihydropteroate synthase (enzyme in malaria parasite important for SP drug resistance) |
| IPTp | Intermittent preventive treatment in pregnancy |
| ITN | Insecticide treated net |
| IRS | Indoor residual spraying |
| LBW | Low birth weight |
| PCR | Polymerase chain reaction |
| RDT | Rapid diagnostic (malaria) test |
| SP | Sulphadoxine-pyrimethamine |

# Introduction

Approximately 31 million pregnant women in malaria-endemic regions in Africa were estimated to be at risk of infection with malaria in 2015;1 no up-to-date estimates are available for regions outside of Africa, but for 2007 the number of malaria exposed pregnancies was estimated to be 94 million.2 Adverse effects in the mother include maternal anaemia and clinical illness, and for the infant effects include prematurity, low birth weight or still birth. Prevention of malaria is recommended, in addition to early detection and treatment.3 Traditionally, malaria has been diagnosed using microscopy; more recent rapid diagnostic tests (RDTs) have been employed. In the context of malaria studies, polymerase chain reaction (PCR) have been introduced; however, this method is generally too expensive and labour-intensive for clinical use in malaria-endemic areas. PCR can detect low-level infections which are not detectable by microscopy or RDTs and thus “subpatent” 4. PCR has enabled the examination of malaria in pregnancy in greater depth, including detection of different species, multiplicity of infection, and molecular markers for drug resistance 5. A review of submicroscopic malaria among non-pregnant persons for mono-infections with *P. falciparum* (106 studies)showed that microscopy detected, on average, 54.1% (95% confidence interval [CI], 50.3–58.2%) of all PCR-detected infections across 106 surveys, whereby the percentage of infections detected by microscopy decreased as transmission level of malaria decreased.4

Among pregnant women it is not clear if subpatent infections cause adverse pregnancy outcomes: some studies detected an association between subpatent infections and adverse pregnancy outcomes,6–8 but not others.9 10 In addition, it is not clear if factors which are known to affect microscopic malaria prevalence in pregnant women, such as gravidity, age, season, HIV infections, and level of malaria transmission have a similar effect on subpatent infections. A systematic review could potentially elucidate the clinical importance of subpatent infections for malaria in pregnancy and what factors may be affecting the relationship between subpatent infections and pregnancy outcome.

## Aims and Objectives of the Study

The aim of this study is to assess the prevalence of and risk factors for subpatent malaria in pregnancy and to assess the association between subpatent malaria and adverse pregnancy outcomes. Our hypotheses are that subpatent malaria is associated with adverse pregnancy outcomes such as maternal anaemia, low birth weight and premature delivery, and that known risk factors for patent malaria, such as low gravidity number, young age, rural setting, high malaria season and HIV infection, are risk factors for subpatent malaria as well. To test these hypotheses, the study is divided into two projects (as listed below):

* Project 1: Evaluation of the association between subpatent malaria and adverse pregnancy outcomes (maternal anaemia/haemoglobin, infant birth weight/low birth weight, infant gestational age at delivery/preterm delivery)
* Project 2: Evaluation of risk factors associated with subpatent malaria in pregnancy
* Project 3: Subpatent malaria during pregnancy over time and effect on pregnancy outcomes

Project 3 will depend on availability of data. Project 2 and 3 will be detailed in a separate analyses plan. The most up-to-date version of the analyses-plans will be available on the website.

# Project 1. Evaluation of the association between subpatent malaria and adverse pregnancy outcomes

Primary Objectives:

* To compare birth weight and low birth weight among women with subpatent and no malaria
* To compare haemoglobin and anaemia among women with subpatent and no malaria during pregnancy and at delivery
* To compare gestational age and preterm delivery among women with subpatent and no malaria

Secondary objectives:

* To compare the outcomes above among women with subpatent malaria with women with patent malaria
* To compare fever among women with subpatent malaria among women with no malaria and patent malaria
* To evaluate the association between subpatent malaria and birth outcome (stillbirth vs. life birth)
* To evaluate the association between subpatent malaria and small-for-gestational age (for datasets which provide birth weight and gestational age assessment)

# Eligibility criteria for inclusion in pooled analysis

We will include datasets with information at either delivery, during pregnancy or both. We will include all pregnant women regardless of parity and gravidity if the dataset has not been selected based on fever or malaria test result. A study will be deemed eligible for the purpose of this analysis (project 1) if they meet the following criteria:

* Cohort study, survey or trial with data on subpatent malaria and one of the outcomes examined during pregnancy (anaemia/haemoglobin) or at the time of delivery (birthweight/LBW, gestational age/preterm delivery, or haemoglobin/anaemia)
* PCR testing or LAMP results and either microscopy or RDT or both available at one or more timepoints (pregnancy/delivery) for one or more compartments (maternal blood, placental blood)

## Essential Data

The dataset must contain the following information:

* Unique patient identifier
* Age or gravidity
* Malaria test by PCR or LAMP in maternal or placental blood
* Malaria test by RDT or microscopy in the same compartment and at the same time as PCR
* One or more of the main outcomes (haemoglobin, birth weight, gestational age)
* Basic information on study design and study site

## Desirable Data (not required for inclusion)

* Gestational age and method
* Enrolment date / Follow up date / Delivery date
* Use of malaria prevention (ITN, IPTp, trial arm, IRS)
* Use of antimalarials during pregnancy
* HIV status
* Fever (history or documented)
* Gender new born
* Outcome delivery (live born or stillbirth)
* Location of living (urban/rural)
* Maternal anthropometry

## Study Exclusion criteria

The following studies will be excluded:

* Case-Reports or Series
* Studies where a selection of the study population is made based on the fever or malaria status of a participant (e.g. studies where PCR is only conducted among women with a positive blood smear or a positive RDT result). If a subsample in a study has been tested by PCR, the study can only be included if the subsample was randomly selected.
* Animal studies, modelling studies

Notes: Studies with only an abstract available will not a-priori be excluded; an attempt may be done to contact the authors to obtain more information if the abstract does not provide enough information for inclusion. Studies where the dataset is not obtained may be included as part of an aggregated analysis.

## Participant exclusion

Participants with no or partially missing information on either PCR or microscopy/RDT will be excluded from all analyses. The number of women for whom the PCR is negative but microscopy or RDT is positive will be reported, but otherwise excluded from the analyses.

# Methodologies

## Data standardisation

On receipt, data will be checked if they belong to the study requested. Data are checked for duplicates, inconsistencies, and unexpected and missing values. For trials we will check the randomization over time, and if groups are balanced for baseline characteristics. The information from the dataset will be checked against the information from available publications with regards to prevalence of malaria and outcomes where possible; if there are discrepancies we will contact the contributors to ask for clarifications. A copy of the data is archived, and all changes will be documented in code (stata).  Variables are recoded and transformed where necessary to ensure a uniform format across datasets. Where possible, outcomes are generated uniformly across all datasets (e.g. anaemia, low birth weight, premature newborn). The datasets will be pooled into a single database of quality-assured individual patient data. We will use the information from the datasets obtained for the analyses.

## Additional data

For each study location, the global positioning system coordinates will be obtained using [Google Earth](https://www.google.com/earth/). Using these coordinates and the midpoint of the study years, the *Plasmodium falciparum* parasite rate in 2–10 year-olds will be obtained from the Malaria Atlas Project as an indicator of [malaria transmission intensity](https://map.ox.ac.uk). This information is available for 2000–2015. For studies before 2000, we will use the estimate for 2000 (PCR was introduced around 1997 for malaria in pregnancy studies). For studies where the HIV-status of the participants is not known, we will obtain the HIV-prevalence among women in the child-bearing age (15–49 years) or, if not available, among the total population for the mid-study year from [UNAIDS](http://aidsinfo.unaids.org). For studies with no information on ITN use or use of IPTp at the time of delivery, we will obtain an estimate using national surveys ([demographic and health surveys](https://dhsprogram.com), [malaria indicator surveys](http://www.malariasurveys.org) or [multiple indicator cluster surveys](http://mics.unicef.org)) closest in time for the administrative region. Because SP-resistance may be a factor important for subpatent malaria, we will evaluate if studies were conducted before or after the introduction of SP in the country; if the study was conducted after the introduction of SP or uses SP as part of a trial arm, we will try to obtain an estimate of the prevalence of SP-resistance markers (*Pfdhps*-A437G and *Pfdhps*-K540E) in the area at the time of study in the area using the [WWARN-SP Molecular Surveyor](http://www.wwarn.org/dhfr-dhps-surveyor/#0) or other sources. Economic status is an additional factor which can affect malaria risk, level of protection (better housing or means for prevention) and impact (access to treatment, nutrition). In the absence of a uniform indicator of socio-economic status across datasets, we will explore if an average level of socio-economic status can be obtained by geographic location and time-period for each study.

# Exposures

For the main exposure the definitions in Table 1 will be used. For the current analyses, two time points will be examined: during pregnancy and at delivery. At delivery, two compartments will be examined: maternal blood and placental blood. Overlap between compartments at the time of delivery will be explored as outcome. Although there is technically also a fourth group when cross-tabulating two blood test results, the group of “microscopy positive, PCR negative” or “RDT positive, PCR negative” will be described for each dataset, and otherwise excluded from further analyses. Generally, this group is small, and may be related to test characteristics (e.g. antibodies for *Plasmodium falciparum* can continue to be detected in the blood by RDT after successful malaria treatment when the DNA of malaria is not detectable, and PCR is negative).11

**Table 1 Definitions used for exposure**

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Patent malaria | Malaria detected by RDT and PCR, or microscopy and PCR. This can be in the maternal blood during pregnancy or at delivery, or in the placental blood at the time of delivery. In addition, patent malaria when present in both compartments at the time of delivery will be explored. |
| Subpatent malaria | Malaria detected by PCR but not by RDT, or malaria detected by PCR but not by microscopy (submicroscopic). This can be in the maternal blood during pregnancy or at delivery, or in the placental blood at the time of delivery. In addition, subpatent malaria when present in both compartments at the time of delivery will be explored. |
| No Malaria | No malaria detected by RDT and PCR, or by microscopy and PCR. At the time of delivery, this will be explored by compartment (placental or maternal blood), or the combination. |

# Outcomes

Primary:

1. Birth weight
2. Low birth weight
3. Gestational age at birth
4. Prematurity
5. Maternal Haemoglobin at birth
6. Maternal: Any Anaemia at birth
7. Maternal: Moderate-to-Severe Anaemia at birth
8. Maternal Haemoglobin during pregnancy
9. Maternal: Any Anaemia during pregnancy
10. Maternal: Moderate-to-Severe Anaemia during pregnancy

Secondary:

1. Maternal fever
2. Stillbirth
3. Small weight for gestational age

The definitions of all the outcomes are available in Table 2. An example schedule of models that will be explored is available for microscopic and submicroscopic malaria in Table 3.

**Table 2: Definitions used for outcomes**

|  |  |
| --- | --- |
| **Outcome** | **Definitions and notes** |
| Low birth weight (LBW) (primary outcome) | A birth weight < 2500 grams  Birth weight should preferably be measured within 24 hours  If weight is not available at birth, weight measured within 7 days will be used and analysis will be adjusted for time of measurement if possible |
| Preterm or premature (PT) (primary outcome) | A gestational age at the time of delivery of <37 weeks  Gestational age should preferably be measured by ultrasound earlier in the pregnancy. However, many methods are in use, e.g. fundal height, first day of the last menstrual period, or a scoring system at the time of delivery. All methods will be accepted and adjusted for in the analysis, if possible. |
| Maternal anaemia (HB11 or HB8) (primary outcome) | Any anaemia is defined as a haemoglobin <11 g/dl. Moderate-to- Severe anaemia is defined as haemoglobin <8 g/dl. If only a haematocrit is available, this will be divided by 3 to get an approximation of the estimate in g/dl12 |
| Fever (data permitting) (secondary outcome) | * History of fever for 1–7 days before the visit * History of fever for an undefined period during pregnancy reported at a visit * Documented fever (≥37.5 °C body temperature) at the visit * Combination of a history of fever or documented fever |
| Small for gestational age (SGA) (secondary outcome) | SGA will be defined as ≤ 10th percentile using the reference values by gestational age and gender of the Intergrowth-21st standard, as reported by Villar et al (2014).13 |
| Stillbirth (SB) (secondary outcome) | Infant born dead at any gestational age as defined by the study involved |

**Table 3: Example of number of analyses for microscopy vs. PCR for primary objectives\***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test result** | **Time point** | **Compartment** | **Reference group** | **Outcomes** |
| Submicroscopic malaria:  -BS negative,  -PCR positive | Pregnancy | Maternal | No malaria by BS and PCR | 1. Haemoglobin |
| 1. Anaemia |
| Malaria by BS and PCR | 1. Haemoglobin |
| 1. Anaemia |
| Delivery | Maternal | No malaria by BS and PCR | 1. Haemoglobin |
| 1. Anaemia |
| 1. Birthweight |
| 1. Low birthweight |
| 1. Gestational age |
| 1. Preterm delivery |
| Malaria by BS and PCR | 1. Haemoglobin |
| 1. Anaemia |
| 1. Birthweight |
| 1. Low birthweight |
| 1. Gestational age |
| 1. Preterm delivery |
| Placental | No malaria by BS and PCR | 1. Haemoglobin |
| 1. Anaemia |
| 1. Birthweight |
| 1. Low birthweight |
| 1. Gestational age |
| 1. Preterm delivery |
| Malaria by BS and PCR | 1. Haemoglobin |
| 1. Anaemia |
| 1. Birthweight |
| 1. Low birthweight |
| 1. Gestational age |
| 1. Preterm delivery |
| Delivery | Maternal and placental blood | No malaria by BS and PCR | 1. Haemoglobin |
| 1. Anaemia |
| 1. Birthweight |
| 1. Low birthweight |
| 1. Gestational age |
| 1. Preterm delivery |
| Malaria by BS and PCR | 1. Haemoglobin |
| 1. Anaemia |
| 1. Birthweight |
| 1. Low birthweight |
| 1. Gestational age |
| 1. Preterm delivery |

Abbreviations: BS: blood smear (microscopy)

\*Same diagram applies to RDT and PCR, or LAMP and microscopy, or LAMP and RDT if there is sufficient data. In addition, combination of results of different compartments will be evaluated.

# Analysis Population

For birthweight and low birth weight analyses, only information from singleton and live births will be used. For gestational age and prematurity, all participants where this information is available and considered reliable is used. Gestational age will be considered not reliable if the proportion premature deliveries in a study is ≥32% (two times the upper limit of the uncertainty interval for the estimated preterm birth rate for sub-Saharan Africa: the estimated preterm birth rate for 2014 for sub-Saharan Africa was 12.0%, 8.6–16.7% uncertainty interval, and for Asia this was 10.4%, 8.7–11.9%).14 If the proportion of premature deliveries is ≥32% in a study, this study will be removed from the analyses for gestational age and preterm delivery. For haemoglobin and anaemia, all participants where this information is available will be used, with separate analyses by time point (pregnancy or delivery).

# Covariates Examined

The following covariates will be extracted for each study (if available) and included in the analyses, as appropriate and described in Table 4.

**Maternal covariates:**

Age

Gravidity

HIV infection

Use of malaria prevention (ITNs, IPTp, IRS).

Antimalarials received in pregnancy

Haematinics received in pregnancy (Haemoglobin or anaemia)

Fever (in the analyses of gestational age or preterm delivery)

Maternal height/weight or mid-upper arm circumference where available (in the analyses of birth weight)

Smoking (birth weight)

Gestational age or trimester (in the analysis of haemoglobin/anaemia in pregnancy)

**Malaria**

Species

Compartment (maternal peripheral blood, placental blood)

**Infant covariates**

Estimated gestational age at delivery

Gender

**Location/ Study site covariates:**

Malaria season at the time of testing

Location of living (urban/rural)

Indicator of malaria transmission intensity for study site

Prevalence of molecular markers of SP resistance in the study area around the time of the study (*Pfdhps*-A437G and *Pfdhps*-K540E)

**Methods**

Method of PCR

Method of LAMP

Method of gestational age assessment

Method of haemoglobin assessment

Method of birthweight assessment

Method of documented fever assessment

**Study level covariates**

Region (America, Asia, West and Central Africa, East and South Africa)

Time period

Study design (trial, cohort, survey)

Risk-of-Bias assessment

# Outline of Statistical Analysis

## Flow chart

A flow chart of the search results will be prepared and an overall study profile as per PRISMA-IPD guidelines of all the studies identified and included in the IPD analyses or as part of the aggregated analyses will be presented

## Baseline characteristics of the participants included in the IPD analysis

Summary of the studies and baseline characteristics of participants included in the analysis will be presented; the summary of studies will include but not be limited to location (country and city/village/district), study design, time period, location of recruitment, sample size, and treatment arms where applicable. External variables, such as an indicator of malaria endemicity, country HIV-prevalence at the time of study, sulfadoxine-pyrimethamine resistance (dhps-K540E, dhps-A437G) and ITN use will be added by location. The summary of participants by study and location will include (as far as available) age, gravidity, use of ITNs, IPTp, antimalarials, season, setting, and HIV status. Outcomes will be summarised by study (haemoglobin/anaemia, birthweight/LBW, gestational age/preterm, fever and birth outcome). The number of available participants will be summarised using percentage or proportion for categorical or binary variables, and median and range for continuous variables.

## Meta-Analysis

Each outcome 1–13 will be analysed following these steps:

* Meta-analyses and forest plots for:

1. Subpatent vs. No malaria
2. Subpatent vs. Patent malaria

We will also do this for the subgroup analyses for the factors listed in Table 4.

* Univariable and multivariable random effects model with malaria status used as a covariate and adjusted for other variables as listed in Table 4.

For continuous outcomes (1,5,8) normal linear regression will be used and for binary outcomes (2,4,6–8,10,12,13) glm or logistic regression will be used. Intercept will be fitted as random effect for study-site, other coefficient will initially be treated as fixed effects.

**In the multivariable model,** variable selection will be conducted following a strategy described in section 8.4.

**Table 4: Analyses in the subpatent malaria in pregnancy IPD study**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Outcome** | **Effect measure** | **Comparator and tests** | **Two stage model (Stata)** | **Subgroup analysis** | **One-stage model (Stata)** |
| Haemoglobin and anaemia in pregnancy | Haemoglobin: Mean difference  Anaemia: Risk ratio or Odds ratio | Exposure:  Submicroscopic malaria  Comparison:  1) No malaria by PCR  2) Malaria by microscopy and PCR  Exposure:  Subpatent malaria  Comparison:  1) No malaria by PCR  2) Malaria by RDT and PCR | IPDmetan   * Regress (continuous outcomes) * Glm (binary outcomes) | * Gravidity * Malaria prevention * Level of malaria transmission * HIV   IPDmetan   * Regress (continuous outcomes) * Glm (binary outcomes) | Xtmixed (continuous outcomes)  Meglm or melogit (binary outcomes)  **Potential Confounders**  Gravidity, age, season, HIV infection, setting (urban or rural), use of malaria prevention (ITNs, IPTp or IRS), use of haematinics, use of antimalarials in pregnancy, gestational age and method or trimester of pregnancy, malaria species, maternal anthropometry  Study level: Haemoglobin method, PCR method, method of gestational age assessment (where not in dataset), level of malaria transmission, HIV (where not available in dataset), study design, quality assessment, region, time period |
| Haemoglobin and anaemia at the time of delivery | Haemoglobin: Mean difference  Anaemia: Risk ratio or Odds ratio | As above | IPDmetan  -Regress (continuous outcomes)  -Glm (binary outcomes) | As above | As above  Gestational age at the time of delivery and method of assessment  In addition: Prevalence of molecular markers of SP resistance in the study area around the time of the study (*Pfdhps*-A437G and *Pfdhps*-K540E) |
| Birthweight  Low birth weight | Birthweight: Mean difference  Low birth weight: Risk ratio or odds ratio | As above | IPDmetan   * Regress (continuous outcomes) * Glm (binary outcomes) | * Gravidity * Malaria prevention * Level of malaria transmission * HIV | Gravidity, age, season, HIV infection, setting (urban or rural), use of malaria prevention (ITNs, IPTp or antimalarials, IRS), use of antimalarials in pregnancy, infant gender, maternal height/weight, smoking, anaemia at delivery  Study level: PCR method, method of gestational age assessment, level of malaria transmission, HIV (where not available in dataset), SP-resistance markers, study design, quality assessment, region, time period |
| Gestational age / preterm delivery | Gestational age: mean difference  Preterm delivery: Risk ratio or odds ratio | As above | IPDmetan   * Regress (continuous outcomes) * Glm (binary outcomes) | * Gravidity * Malaria prevention * Level of malaria transmission * HIV | Gravidity, age, season, HIV infection, setting (urban or rural), use of malaria prevention (ITNs, IPTp or antimalarials, IRS), use of antimalarials in pregnancy, fever, infant gender  Study level: PCR method, method of gestational age assessment (where not in dataset), level of malaria transmission, HIV (where not available in dataset), SP-resistance markers, study design, quality assessment, region, time period |

# Statistical Methodology

## Flow chart

The results of the search, eligibility, contacts and inclusion and exclusion will be documented in the flow chart as provided by [Prisma for IPD](http://prisma-statement.org/Extensions/IndividualPatientData.aspx). A characteristics table of participating studies by type of involvement (in IPD analysis or in aggregated analysis) will be compiled; this will include information on study, study type, country of conduct of study, period of study, number of participants and exposures, outcomes and covariates available.

## Descriptive statistics

Descriptive statistics will use mean and variance/standard deviation if the data are normally distributed, geometric mean and range if data are log-normally distributed or median and range/ interquartile range otherwise. Statistical tests or plots will be used to check the assumption of normality. The summary statistics will be further broken down by age and gravidity.

## Analytical Methods

For all analyses the following steps will be conducted:

*Two-stage approach*

1. Outcomes will be generated from raw data. From the compiled dataset we will assess the number of studies with exposures for a certain outcome. Using IPDmetan (Stata version 14.2) a forest plot will be constructed to visualise the outcome by exposure for each study. This allows to assess if there are anomalies in the data. The pooled estimate will be derived using the DerSimonian and Laird random effects method. Covariates will not yet be included in the models.

*One stage-approach*

1. In the second phase, we will use a one-stage approach with regression analysis. For continuous variables such as haemoglobin, birthweight, and gestational age, we will use xtmixed and place a random effect on the intercept, to allow for heterogeneity in baseline risk across studies. For binary variables, meglm will be used with log as link function to obtain risk ratios. Alternative models such as melogit will be explored if convergence issues emerge. We will visualise results of the one stage approach using the command IPDforest in Stata for xtmixed and melogit.

## Assessment of heterogeneity between sites

We will evaluate the standard deviation of the random effect as the measure of heterogeneity in mixed level models.

## Model selection for determinants

Model building will be carried out first by investigating if any of the available co-variates (Table 4) are related to the outcome using the regression model. We will include the co-variates one at a time as a fixed effect to assess if there is a significant association between the co-variate and the outcome (univariate analysis). All significant variables will be re-introduced (at the same time) with our exposure of interest, and non-significant factors will be removed. The exposure of interest will be kept in the model, even if not significant. When the final model with only significant factors (and exposure of interest) has been identified, all factors that were non-significant in the univariate analysis, will be re-introduced one by one, to assess the association again. Co-variates in the final model will be examined when introduced as a fixed or a random-effect, to assess if there is between study difference. We will assume that the effects of the covariates will be the same across all sites, except for the following covariates age (young age: < 20 yrs vs. older) and gravidity (paucigravidae vs. gravidae 3+) for which interaction terms with endemicity will be tested as the impact of gravidity (i.e. pregnancy-specific immunity) is known to be different depending on the endemicity.

Any known confounding factors will be forced into the model even if they are statistically non-significant (e.g. gravidity for birthweight). If data permit, linear and non-linear relationship will be examined for continuous variables. Colinearity will be examined between age and gravidity, before including the models.

## Missing data

There are different types of missing data:

1) the variable may not have been collected;

2) the variable was collected but incomplete;

3) the variable was collected but not included in the dataset.

In the last case, we will contact the original authors of the study to obtain the relevant missing data. If the variable of interest was not collected, we will try to assess if we can obtain information from external sources (e.g. ITN use, HIV-prevalence). We will examine if the non-availability of a co-variate is associated with characteristics of a study (e.g. malaria endemicity), to assess if this is a potential source of bias. To assess the impact of missing co-variates, sensitivity analysis will be performed to see if our main conclusion is affected or not by the exclusion of participants with missing data. For partial missing data, assuming that the data are missing at random, we will explore a multiple imputation approach to handle missing data using one of the several commonly used statistical packages i.e. R and Stata. The number of imputed datasets will be decided on the basis of the fraction of missing information. E.g. if the missing data is <20%, imputation may be conducted, and the analysis repeated.

## Sensitivity Analysis

Several methods will be used to assess the reliability of the results. Methodology of e.g. PCR, gestational age, and haemoglobin are used as covariates, in addition to study design, and location of recruitment. A risk-of-bias (quality) assessment will be conducted among all studies, as explained in the protocol, and studies will be assessed as higher, lower or unclear risk of bias. This factor will be examined in the models. Other approaches include:

1. Removal of studies one by one and rerun final models, to see if some studies affect the covariates effects more than others.
2. Missing co-variates: we will assess the results by availability of co-variates and examine if missing co-variates are at random or associated with certain study characteristics such as malaria endemicity.
3. Missing data (incomplete information for a variable): For the final model, multiple imputations will be conducted for missing data, and results compared with and without imputation.
4. Aggregated data extracted from the publications for studies not included in the IPD meta-analysis will be added to the forest plots to explore bias between studies with- and without datasets, using the “ad” option in ipdmetan command in stata. Pooled estimates for IPD data and aggregated data will be presented together with the overall estimate of pooling all data together.

## Risk of bias

There is a risk of bias if studies which are not included in the IPD are different from the studies included. This will be explored by comparing characteristics between the groups, and by including the study results of studies without datasets as aggregated data.

Women with missing outcomes may be different from women with outcomes available. We will assess for each study the proportion and characteristics of women without the outcomes of interest to assess if this might be a source of bias. If the proportion of women without the outcome is large (10% or more) we will compare the characteristics with women with the outcome available and record explanatory factors.

# Tools

All statistical analyses will be carried out using R 3.1.2 (released on 2014-10-31 by The R Foundation for Statistical Computing) and Stata 14.2 (StataCorp, 4905 Lakeway Drive, College Station, Texas 77845 USA). Using alternative statistical software will not require amendment of this plan.

# References

1. Walker PG, Floyd J, ter KF, et al. Estimated impact on birth weight of scaling up intermittent preventive treatment of malaria in pregnancy given sulphadoxine-pyrimethamine resistance in Africa: A mathematical model. *PLoS Medicine* 2017;14:e1002243. doi: 10.1371/journal.pmed.1002243 [doi];PMEDICINE-D-15-03688 [pii]

2. Dellicour S, Tatem AJ, Guerra CA, et al. Quantifying the number of pregnancies at risk of malaria in 2007: a demographic study. *PLoS Medicine* 2010;7:e1000221. doi: 10.1371/journal.pmed.1000221 [doi]

3. Desai M, Dellicour S. Effects of malaria and its treatment in early pregnancy. *Lancet Infectious Diseases* 2012;12:359-60. doi: S1473-3099(11)70345-0 [pii];10.1016/S1473-3099(11)70345-0 [doi]

4. Okell LC, Ghani AC, Lyons E, et al. Submicroscopic infection in Plasmodium falciparum-endemic populations: a systematic review and meta-analysis. *J Infect Dis* 2009;200:1509-17. doi: 10.1086/644781

5. Arango E, Maestre A, Carmona-Fonseca J. Effect of submicroscopic or polyclonal Plasmodium falciparum infection on mother and gestation product: systematic review. *Revista Brasileira de Epidemiologia* 2010;13:373-86.

6. Mockenhaupt FP, Rong B, Till H, et al. Submicroscopic Plasmodium falciparum infections in pregnancy in Ghana. *Tropical Medicine and International Health* 2000;5:167-73.

7. Adegnika AA, Verweij JJ, Agnandji ST, et al. Microscopic and sub-microscopic Plasmodium falciparum infection, but not inflammation caused by infection, is associated with low birth weight. *American Journal of Tropical Medicine and Hygiene* 2006;75:798-803.

8. Cottrell G, Moussiliou A, Luty AJ, et al. Submicroscopic Plasmodium falciparum infections are associated with maternal anemia, premature births, and low birth weight. *Clinical Infectious Diseases* 2015;60:1481-8.

9. Mockenhaupt FP, Bedu-Addo G, von Gaertner C, et al. Detection and clinical manifestation of placental malaria in southern Ghana. *Malaria Journal* 2006;5:119.

10. Walker-Abbey A, Djokam RR, Eno A, et al. Malaria in pregnant Cameroonian women: the effect of age and gravidity on submicroscopic and mixed-species infections and multiple parasite genotypes. *American Journal of Tropical Medicine and Hygiene* 2005;72:229-35.

11. Dalrymple U, Arambepola R, Gething PW, et al. How long do rapid diagnostic tests remain positive after anti-malarial treatment? *Malar J* 2018;17:228. doi: 10.1186/s12936-018-2371-9

12. Carneiro IA, Drakeley CJ, Owusu-Agyei S, et al. Haemoglobin and haematocrit: is the threefold conversion valid for assessing anaemia in malaria-endemic settings? *Malar J* 2007;6:67. doi: 10.1186/1475-2875-6-67

13. Villar J, Cheikh Ismail L, Victora CG, et al. International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. *Lancet* 2014;384:857-68. doi: 10.1016/S0140-6736(14)60932-6

14. Chawanpaiboon S, Vogel JP, Moller AB, et al. Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. *Lancet Glob Health* 2019;7:e37-e46. doi: 10.1016/S2214-109X(18)30451-0