Statistical Analysis Plan
WWARN Vivax Haematology study group: A pooled analysis of haematological recovery after treatment of *Plasmodium vivax*
Version 0.1

WorldWide Antimalarial Resistance Network (WWARN)

Suggested citation: Statistical Analysis Plan, WWARN Vivax Haematology study group: A pooled analysis of haematological recovery after treatment of *Plasmodium vivax*

Version History

Version number	Revision(s) & reason for amendment	Release date
V.1		

WorldWide Antimalarial Resistance Network (WWARN)
www.wwarn.org

1. Contents

2. Introduction and Rationale

A significant proportion of the morbidity and mortality attributed to *P. vivax* is a consequence of severe anaemia. With each recurrence of *P. vivax* parasitaemia there is a cumulative risk of anaemia (1). Malaria-associated anaemia is a complex phenomenon, related to increased red cell destruction and haemopoeitic suppression, compounded by nutritional status, helminth carriage and drug induced haemolysis. The degree of anaemia caused by vivax malaria and its risk factors has not been evaluated widely [1].

Control and ultimate elimination of *P. vivax* requires the safe and effective prevention of recurrent infections and this requires radical cure targeting both blood and liver stages of the parasite. However, the only widely available drug that is active against the liver stages is primaquine, an 8-aminoquinoline drug that can cause drug-induced haemolysis particularly in individuals with glucose-6-phosphate dehydrogenase (G6PD) deficiency. In order to quantify the risks and benefits of *P. vivax*

radical cure, it is crucial to determine the normal haematological response following *P. vivax* infection and treatment.

3. Aim of the study

The aim of this study is to investigate the acute haematological effect of *Plasmodium vivax* parasitaemia and its subsequent recovery following treatment.

4. Eligibility criteria

4.3.1 *Essential inclusion criteria*

- Prospective clinical efficacy studies of uncomplicated vivax monoinfection
- Asexual parasitaemia at enrolment
- Haemoglobin (hb) or haematocrit (hct) at enrolment
- Study meta-data as described in the Clinical Data Management and Statistical Analysis Plan (2)
- Baseline data on patient age and gender

4.3.2 *Desirable criteria*

- Information on the timing and dose of drug administration.
- Exact mg/kg dosing of schizontocidal and primaquine administration
- Weight of the patient
- Information on splenomegaly, hepatomegaly
- Malnutrition as gauged by weight and age or height, or Middle Upper Arm Circumference (MUAC)
- Qualitative or quantitative assessment of G6PD status
- Parasite density at days 1, 2, 3
- Outcome of malaria treatment according to standardised WHO criteria (3)
- Information on previous vivax episodes

4.3.3 *Exclusion criteria*

- Pregnancy

5. Data Pooling

A systematic review of all prospective clinical efficacy trials involving *Plasmodium vivax* mono-infection has been completed. Trials undertaken since the year 2000 that fulfil the study criteria have been targeted through direct email to the corresponding author and/or principal investigator. Data from unpublished and ongoing clinical studies will also be included if available. Once data are uploaded into the WWARN repository, they will be curated and standardized using the WWARN Data Management and Statistical Analysis Plans (2) for clinical data and pooled into a single database of quality-assured individual patient data.

6. Outline of Statistical Analysis

6.1. Specific objectives of the study

- To quantify the acute reduction in haemoglobin associated with acute vivax malaria before and after treatment
- To identify risk factors associated with anaemia at the start of treatment for acute vivax malaria.
- To identify risk factors associated with the development of anaemia during the follow-up phase of vivax malaria (up to 28 or 42 days), with and without primaquine
- To estimate the time to recovery from anaemia after administration of chloroquine and compare this with non-chloroquine treatments
- To quantify the effect of primaquine mg/kg dose on haemoglobin reduction and time to anaemia recovery.

6.2. Study endpoints

Overall haematological profile:

- Mean haemoglobin at days 0, 1, 2, 3, 7, 14, 28, 35 and 42
- Predicted normal haemoglobin level and time to predicted normal haemoglobin level

At Enrolment:

1. Mean haemoglobin
2. Presence of anaemia (Hb<10g/dL) or severe anaemia (Hb<7g/dL)
3. Fractional or absolute reduction in haemoglobin at enrolment evaluated against the patients' predicted normal haemoglobin level (NHL) in children and adults

After treatment

4. Risk of anaemia or severe anaemia within 28 days / 42 days
5. Time to haemoglobin nadir
6. Maximal fractional or absolute change in haemoglobin from baseline
7. Fractional change or absolute change in haemoglobin at baseline compared to the NHL (Malaria attributable fall before treatment)
8. Maximal fractional change or absolute change in haemoglobin at the nadir compared to the NHL (Total malaria attributable fall in haematocrit).
9. Time to haematological recovery

6.3 Definitions of Endpoints

Absolute reduction in haemoglobin between times t_1 and t_2 will be defined as $hb(t_2)- hb(t_1)$.

Fractional reduction in haemoglobin between times t_1 and t_2 will be defined as $(hb(t_2)- hb(t_1))/hb(t_1)$, or where $hb(t_i)$ denotes measured or estimated haemoglobin at time t_i.

Normal haemoglobin level (NHL) will be derived from data from this pooled analysis based on the shape of population haemoglobin curves over time for age, sex and region. Only data from patients who do not have a recurrence before day 42 will be included.

Duration of haematological recovery will be defined as the time from enrolment to the first time of reaching haemoglobin ≥10 g/dl after any documentation of anaemia within the first 7 days of presentation.

Time of haemoglobin nadir will be defined as the time when the minimum haemoglobin was recorded, provided measurements were available on all days 0, 3, 7, 14, 21 and 28.

6.4. Study and patient characteristics

The following baseline characteristics will be included in the analysis:

Site: transmission intensity, regional relapse periodicity, chloroquine resistance

Patient: age, sex, weight, nutritional status, history of malaria within 28 days, history of fever within the last 24 hours, fever (>37.5^{0}C), G6PD status

Drug: schizontocidal treatment and mg/kg dose, primaquine treatment, timing and mg/kg dose, supervision of drug intake (full or partial), co-administration with fat, dose vomiting

Laboratory: baseline parasitaemia, baseline gametocytes

Children will be considered as aged <14 years with childhood age stratified into < 1 years, 1 to 4 years, 5 to 11 years, and 12 to 14 years.

The nutritional status of children aged <5 years of age will be calculated as a weight-for-age z-score, using the igrowup package developed by WHO (4). Those with weight-for-age z-scores < -2 (i.e. below the 3rd centile) will be classified as underweight-for-age (termed underweight).

Treatment will be classified as supervised if all doses were directly observed, partially supervised if at least the 3 morning doses were observed, and not-supervised if fewer doses were observed.

Total doses will be calculated from the recorded number of tablets administered per dose if this information is available in the individual patient data. If no individual patient dosing data was available, dose will be estimated using the protocol dosing schedule.

In studies with haematocrit measured instead of haemoglobin, haematocrit will be converted to haemoglobin using the following relationship (5):

$$Haematocrit\ (ht) = 5.62 + 2.60 * Haemoglobin\ (hb)$$

Anaemia will be defined according to WHO guidelines (6), which includes age stratification.

WHO definitions of efficacy outcome will be used (3).

Patients early parasitological response will also be evaluated in the form of (a) estimated parasite clearance half life using the WWARN PCE tool (7); (b) positivity on Day 1; (c) positivity on Day 2.

For each study, locations of study sites will be recorded. Each location will be categorised into:

a) *low, moderate and high transmission settings* based on the observed study site reinfection rate , and the malaria endemicity estimates obtained for study sites and year from the Malaria Atlas Project (8). *P. vivax* Parasite Rate (PvPR) < 0.15 will be categorized as "low" transmission areas, PvPR ≥ 0.15 & < 0.40 will be classified as "moderate" transmission areas, and PvPR ≥ 0.40 will be classified as "high" transmission areas.
b) *low, medium and high periodicity of relapses* according to Battle's regions (9).

6.5 Summary of statistical analyses

Descriptive statistics and baseline characteristics of study sample:

1. A summary (study profile) of the relevant trials uploaded to the WWARN repository will be presented to highlight potential selection bias.
2. A summary of the relevant studies will be presented, including (but not restricted to) treatment tested, inclusion and exclusion criteria, follow up duration, study populations, parasitaemia sampling scheme, use of haematinic agents and method of haemoglobin testing.
3. The baseline characteristics of the eligible studies will be described by country, transmission site(s), regional relapse periodicity, chloroquine resistance and treatment regimens. Tests of statistical significance will not be undertaken for baseline characteristics; rather the clinical importance of any differences in the baseline distributions will be noted.

   The distribution of continuous variables (e.g. haemoglobin, age, parasitaemia) will be described using the mean and standard deviation if the data are normally distributed, geometric mean and 95% reference range if the data are normally distributed following a log transformation, or the median and interquartile range if the data are non-normally distributed.

Description of overall haematological profiles:

4. Figure of mean (+/- 2 sd) haemoglobin concentrations at each available time point following treatment will be presented by: transmission intensity, relapse periodicity, treatment group (including primaquine use).
5. Separate analyses will be performed for gender as well as children and adults with children also stratified by age category.

6. Normal haemoglobin levels

At Enrolment (before treatment):

7. Mean haemoglobin at enrolment

Univariable and multivariable linear regression for the haemoglobin on enrolment will be performed, with a random intercept for study-site. The following covariates will be examined: age, sex, weight, nutritional status, G6PD status, history of malaria within the last 28 days, history of fever, baseline parasitaemia, presence of gametocytes on enrolment, transmisison intensity and relapse periodicity. Model selection will be undertaken as described in Annex 2.

The linearity of the association between baseline haemoglobin and the continuous covariates; age, baseline parasitaemia, transmission intensity and relapse periodicity, will be assessed visually by superimposing a lowess curve on the scatterplot and statistically using fractional polynomials.

8. Anaemia at enrolment

Univariable and multivariable logistic regression of risk factors for anaemia on enrolment will be performed, with random effects for study-site and model building as described in Annex 2. Covariates to examine will include: age, sex, weight, nutritional status, G6PD status, history of malaria within the last 28 days, history of fever, baseline parasitaemia, presence of gametocytes on enrolment, transmisison intensity and relapse periodicity.

9. Fractional or absolute reduction in haemoglobin at enrolment evaluated against the patients predicted normal haemoglobin level (NHL)

Linear regression for fractional (absolute) reduction in haemoglobin on enrolment will be performed, with a random intercept for study-site and model building as described in Annex 2. The following covariates will be examined: age, sex, weight, nutritional status, G6PD status, history of malaria within the last 28 days, history of fever, baseline parasitaemia, presence of gametocytes on enrolment, transmisison intensity and relapse periodicity. Depending on the distribution of the outcome variable, appropriate transformation will be used to ensure residual are normally distributed.

During Follow Up (After Treatment):

10. Risk of anaemia within 28 days / 42 days

Survival analysis for time to anaemia during follow-up (28 or 42 days) will be performed, with random effects for study-site. Covariates to examine will include: age, sex, weight, nutritional status, G6PD status, history of malaria within the past 28 days, history of fever, baseline parasitaemia, prevalence of parasitaemia on days 1 and 2, transmission intensity, relapse periodicity, baseline haemoglobin, presence of gametocyte on enrolment, schizontocidal treatment, primaquine treatment and timing, and recurrent infection. Only patients without anaemia on enrollment will be included. Model building for a prediction model of time to anaemia is described in Annex 2.

11. Fractional or absolute reduction in haemoglobin during follow-up evaluated against the baseline haemoglobin and against the NHL

Linear regression for fractional (absolute) reduction in haemoglobin during follow-up will be performed, with a random intercept for study-site and model building as described in Annex 2. The

following covariates will be examined: age, sex, weight, nutritional status, G6PD status,history of malaria within the past 28 days, history of fever, baseline parasitaemia, prevalence of parasitaemia on days 1 and 2, transmisison intensity, relapse periodicity, baseline haemoglobin, presence of gametocyte on enrolment, schizontocidal treatment, primaquine treatment and timing, and recurrent infection. Depending on distribution of outcome variable, appropriate transformation will be used to ensure residuals are normally distributed.

12. Maximum Fractional fall in haemoglobin after treatment evaluated against the baseline haemoglobin and against the NHL

Linear regression for maximum fractional reduction in haemoglobin from enrolment and from predicted NHL during follow-up will be performed, with a random intercept for study-site and model building as described in Annex 2. The following covariates will be examined: age, sex, weight, nutritional status, G6PD status,history of malaria within the past 28 days, history of fever, baseline parasitaemia, prevalence of parasitaemia on days 1 and 2, transmisison intensity, relapse periodicity, baseline haemoglobin, presence of gametocyte on enrolment, schizontocidal treatment, primaquine treatment and timing, and recurrent infection. Depending on distribution of outcome variable, appropriate transformation will be used to ensure residuals are normally distributed. If this is not possible the outcome variable may be categorised and alternative (logistic, ordered logistic) models will be considered.

Figure of mean (95% confidence interval) fractional change in haemoglobin from enrolment during follow-up will be presented according to treatment group, age categories, transmission intensity, baseline parasitaemia, baseline haemoglobin, day 1 clearance and day 2 clearance.

Figure of mean (95% confidence interval) fractional change in haemoglobin from the predicted NHL at enrolment and after treatment will be presented according to treatment group, age categories, transmission intensity, baseline parasitaemia, baseline haemoglobin, day 1 clearance and day 2 clearance.

Figure of mean (95% confidence interval) fractional change in haemoglobin from the predicted NHL at enrolment and after treatment will be presented according to mg/kg dose of chloroquine, other schizontocidal agents (if numbers allow) and primaquine.

13. Day of nadir in haemoglobin

Day of nadir will be assessed visually by superimposing a lowess curve on the scatterplot of haemoglobin vs time, for subgroups of patients with and without day 3 and day 7 data.

14. Time to haematological recovery

Haematological recovery will be defined in a subgroup of patients with clinically relevant anaemia (Hb<10 g/dl) at enrolment or within the first week of treatment, and the first occurrence of Hb above this level. Survival analysis for time to haematological recovery will be performed, with random effects for study-site. Covariates to examine in will include: age, sex, weight, nutritional status, history of malaria within the past 28 days, history of fever, baseline parasitaemia, prevalence of parasitaemia on days 1 and 2, transmisison intensity, relapse periodicity, baseline haemoglobin, presence of gametocyte on enrolment, schizontocidal treatment, primaquine treatment and timing, G6PD status and recurrent infection.

Model building to create a prediction model of time to haematological recovery will be performed as described in Annex 2.

7 Tools

All statistical analyses will be carried out using Stata version 13.0 and R 3.1.0 released on 2014-04-10 by The R Foundation for Statistical Computing. However, when equivalent statistical methods are applied, changing the use of statistical software does not require amendment of this SAP.

8 Study Group Governance, Management, Coordination and Publication Policy

The Vivax Anaemia Study Group comprises participating investigators who contribute relevant data sets to the pooled analysis. Data sets will remain the property of the investigator and will not be shared without their consent. The WWARN statistician(s) will oversee the statistical analyses. Participating investigators will be recognised in publication as contributors under the banner of the **Vivax Anaemia Study Group**. A Writing Committee will coordinate activities including data analysis, and drafting of publications and reports for complete group review. The Writing Committee will comprise Ric Price, Rob Commons, Julie Simpson, the WWARN statisticians Kasia Stepniewska and Rashid Masoor and other interested investigators. They are responsible for undertaking the data analysis and preparation of the manuscript. Authors will be recognized according to the ICMJE guidelines and the WWARN publication policy (10).

9 Potential Policy Outcome

The data provided by this analysis will be used to inform policy makers and research on the relative risks of anaemia, and the comparative risks and benefits of alternative treatment options including primaquine for radical cure.

10 References

1. Douglas NM, Lampah DA, Kenangalem E, Simpson JA, Poespoprodjo JR, Sugiarto P, et al. Major burden of severe anemia from non-falciparum malaria species in Southern Papua: a hospital-based surveillance study. PLoS Med. 2013;10(12):e1001575; discussion e.
2. WorldWide Antimalarial Resistance Network. Data Management and Statistical Analysis Plan v1.2. Oxford; 2012.
3. World Health Organization. Methods for surveillance of antimalarial drug efficacy. Geneva: World Health Organization; 2009.
4. World Health Organization. WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development. Geneva; 2006.

5. Lee SJ, Stepniewska K, Anstey N, Ashley E, Barnes K, Binh TQ, et al. The relationship between the haemoglobin concentration and the haematocrit in Plasmodium falciparum malaria. Malar J. 2008;7:149.
6. World Health Organization. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity Geneva: World Health Organization; 2011 [Available from: http://www.who.int/vmnis/indicators/haemoglobin.pdf.
7. Flegg JA, Guerin PJ, White NJ, Stepniewska K. Standardizing the measurement of parasite clearance in falciparum malaria: the parasite clearance estimator. Malar J. 2011;10:339.
8. Gething PW, Elyazar IR, Moyes CL, Smith DL, Battle KE, Guerra CA, et al. A long neglected world malaria map: Plasmodium vivax endemicity in 2010. PLoS Negl Trop Dis. 2012;6(9):e1814.
9. Battle KE, Karhunen MS, Bhatt S, Gething PW, Howes RE, Golding N, et al. Geographical variation in Plasmodium vivax relapse. Malar J. 2014;13:144.
10. WorldWide Antimalarial Resistance Network. WWARN Publication Policy Oxford, UK: WWARN; 2015 [Available from: http://www.wwarn.org/sites/default/files/attachments/documents/wwarn_publication_policy.pdf.
11. WorldWide Antimalarial Resistance Network. Vivax Surveyor: WWARN; 2016 [Available from: 31 July 2016.
12. Collett D. Modelling Survival Data in Medical Research. Second edition ed: Boca Raton: Chapman and Hall; 2003.

11 Annex 1

A.1 List of available covariates

Description	Type
Patent parasitaemia on day1, day2 and day3	Primary Response
Parasite Clearance on day 1 and day 2	Secondary Response
History of Fever (0/1) at inclusion within 24 hours	Baseline Variable
Haemoglobin at inclusion	Baseline Variable
Vivax density at Inclusion	Baseline Variable
Gam (/µL) at inclusion	Baseline Variable
Max Temp Day0	Available Variable
D0 Ht<20%	Baseline Variable
Age in Years	Baseline Variable
Gender	Baseline Variable
Weight	Available Variable
G6PD status	Available Variable
Antimalarial in last 28 days	Available Variable
Transmission intensity	Available Variable
Relapse periodicity region	Available Variable
Max Vivax Asexual parasitaemia on Day1	Available Variable
Max Vivax Asexual parasitaemia on Day2	Available Variable
Max Vivax Asexual parasitaemia on Day3	Available Variable

Max Temp Day1	Available Variable
Max Temp Day2	Available Variable
Max Temp Day3	Available Variable
Dosing method (single day, broken down over days etc.)	Available Variable
Drug timing (eg primaquine)	Available variable
Total mg/kg dose at each day of dosing regimen	Available Variable
Total mg/kg dose during course	Available Variable

12. Annex 2 - Model selection for determinants

Linear regression and logistic regression models

Model building will be carried out first by investigating if any of the available variables (Annex 1) are related to the outcome using linear regression for continuous outcomes and logistic regression for categorical outcomes.

Initially all possible risk factors will be examined in the univariable model, and will be included in model building in the multivariable analysis. Known confounders will be fitted first and forced into the model. Variables and covariates will then be added in a stepwise forward fashion using model deviance/Likelihood Ratio Test (LRT) i.e. changes in log likelihood $(-2\ Log\hat{L})$ will be compared (for nested models) to identify the variables which results in a significant reduction in $-2\ Log\hat{L}$ at 5% level of significance. Akaike's Information Criterion (AIC) will be used to compare competing non-nested models; models with smaller AIC will be preferred.

Sensitivity analysis will be carried out by removing one study site at a time. Additional sensitivity analysis will be undertaken to assess the effect of chloroquine resistance using studies undertaken in areas *with and without evidence of chloroquine resistance as defined by the* WWARN Vivax Surveyor (11).

Cox regression models

To create a prediction model of time to haematological event a Cox regression model and models with parametric hazard functions such as: Gompertz, Weibull, lognormal and log-logistic will be examined and the best regression model will be selected based on Cox-Snell residuals (12). Any known confounding factors (age, gender, region) will be forced into the multivariate model even if they are statistically non-significant. All the variables which were significant in the univariable analysis at 10% level of significance will be kept for the multivariable analysis. Inclusion of covariates in the final model will be determined based on how they improve the overall model (likelihood ratio test) and if they change the coefficient estimates for other factors and based on the residuals. Robustnes of the coefficients in the final model will be explored using 1000 boostrap samples. Coefficient of Variation (CV, %) of the estimates derived from bootstrap analysis will be reported.

Sensitivity analysis will be carried out by removing one study site at a time and coefficient of variation (CV) around the parameter estimates will be presented. Sensitivity analysis will also be undertaken to assess the effect of chloroquine resistance using studies undertaken in areas *with and without evidence of chloroquine resistance as defined by the* WWARN Vivax Surveyor (11).