**Suggested citation:** Statistical Analysis Plan, Gametocyte Carriage Study Group: Investigating factors effecting gametocyte carriage

**Version History**

<table>
<thead>
<tr>
<th>Version number</th>
<th>Revision(s) &amp; reason for amendment</th>
<th>Release date</th>
</tr>
</thead>
<tbody>
<tr>
<td>v1.0</td>
<td></td>
<td>04.12.13</td>
</tr>
<tr>
<td>V2.0</td>
<td></td>
<td>08.04.14</td>
</tr>
<tr>
<td>V2.1</td>
<td></td>
<td>28.04.14</td>
</tr>
<tr>
<td>V2.2</td>
<td></td>
<td>08.05.14</td>
</tr>
<tr>
<td>V2.3</td>
<td></td>
<td>28.05.14</td>
</tr>
</tbody>
</table>

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

The WHO recommends that patients with uncomplicated malaria be treated with a combination of two unrelated drugs, with the currently preferred option being Artemisinin Combination Therapies (ACTs). Unlike most schizontocidal agents the artemisinin derivatives have activity against the stage II and IV gametocyte asexual stages, a property that has potential to reduce the transmission potential of malaria and decrease the spread of drug resistant parasites. However the consequences of this gametocytocidal activity vary depending on the drug regimen, the doses administered, the local prevalence of drug resistance and host immunity (Arinaitwe et al. 2009; Price et al. 1996; Bassat et al. 2009; Kamya et al. 2007; Mens et al. 2008; Nambozi et al. 2011; Yeka et al. 2008; Zongo et al. 2007). A variety of factors are associated with gametocyte carriage including; presentation with fever, a high asexual parasitaemia, age, anaemia, pure infection of P.f and a palpable spleen (Akim et al. n.d.; J. T. Bousema et al. 2004; Price et al. 1999; Sowunmi et al. 2011; Stepniewska et al. 2008). The heterogeneity of these factors confounds comparison between studies. Furthermore the metrics for measuring gametocyte carriage vary considerably from simple proportions observed during follow-up, to measures of incidence density and ultimately to the infectivity of the patients’ blood to mosquitoes (T. Bousema and Drakeley 2011; Stepniewska et al. 2008). Whilst there is good evidence that ACTs are associated with a significant reduction in gametocyte carriage, the crucial factor is the ability of patients to infect the mosquito vector.
In the era of malaria elimination, the public health benefit of different ACTs and their ability to reduce malaria transmission are critical considerations in guiding antimalarial policy. There is an urgent need to review gametocyte carriage across a variety of settings using standardised analytical approaches, to identify the key determinants underlying ACTs role in reducing transmission potential. To achieve this requires a standardised database and a set of metrics which are derived in a systemic manner. Data within the WWARN repository provide an excellent opportunity to achieve this.

2. Aim of the study
The purpose of this study group is to assess the risk factors associated with gametocyte carriage and clearance across a range of endemic settings and drug treatments.

3. Eligibility criteria for inclusion in pooled analysis
Studies on uncomplicated *P. falciparum* malaria with patients receiving any antimalarial treatment will be considered for inclusion in this pooled analysis.

A study will be deemed eligible for the purpose of this analysis if they meet the following criteria:
- Uncomplicated *P. falciparum* malaria (either alone or mixed)
- Asexual and sexual parasite counts at day 0
- Known method of counting parasites/gametocytes (identify thick or thin film origin too)

4. Methods for data standardization
The data sets uploaded to the WWARN repository will be standardized using the WWARN Data Management and Statistical Analysis Plans (DMSAP v1.1) for clinical data and pooled into a single database of quality-assured individual patient data. Data will remain the property of the individual donor(s) and publication will be under a ‘study group’ authorship with individual contributors still listed for the purposes of PubMed searching.

1. [http://www.wwarn.org/partnerships/data/methodology/clinical](http://www.wwarn.org/partnerships/data/methodology/clinical)
5. **Study Objectives**

In this pooled analysis we will seek to investigate the following:

1. Gametocyte carriage at different follow-up days, by ACT treatment
2. Risk factors for gametocyte carriage before initiation of treatment
3. Risk factors for appearance of gametocytes after treatment with ACT (d7 onwards)
4. Factors associated with duration of gametocyte carriage
5. Association between recrudescence and gametocyte carriage
6. The relationship between the speed of asexual parasite clearance and gametocyte clearance and carriage
7. Association between total parasite load and total gametocyte load

6. **Study endpoints**

**Primary Endpoints:**

The presence of gametocytes on enrolment (within first 24hrs)

The carriage of gametocytes during follow-up

**Secondary Endpoints:**

Prevalence of gametocytes at follow-up days

Rate of gametocyte clearance

Duration of gametocyte carriage within follow-up

Maximum gametocyte density

Area under gametocyte curve within study follow-up

Area under the gametocyte curve normalized by the maximum gametocytaemia

Area under the infectivity curve within study follow-up

6.1 **Definitions**

**Gametocytes on enrollment** is be defined as any sexual parasitaemia count/presence within 24hrs of the reading, in patients in whom this was assessed by thick film examination.

**Gametocyte carriage during follow-up** is defined as patent gametocytaemia after enrollment (>24hrs) up to study follow-up, whilst taking account of reinfection rates, transmission levels, and concurrent asexual parasitaemia results within patients.
The appearance of gametocytes will be defined as gametocyte carriage during study follow-up in patients with no detectable gametocytes present at enrollment (within first 24hrs).

Prevalence of gametocytes during follow-up will be determined on days 7, 14, 21 and 28-42 according to patent gametocytaemia by microscopy on each day of observation. Patients with missing counts on that day will be excluded from the analysis, unless a missing count is between two positive counts (it will be assumed to be positive).

The duration of gametocyte carriage during follow-up is defined as the time between the first positive gametocytaemia record and half-way between the last positive record and the subsequent negative record (e.g. last positive record +3.5days). For patients with missing records after a positive record, the duration of gametocytes carriage is censored at the day of the last record +3.5days. Some examples are given below.

<table>
<thead>
<tr>
<th>Day</th>
<th>0</th>
<th>7</th>
<th>14</th>
<th>21</th>
<th>28</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pt1</td>
<td>+</td>
<td>null</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>Gct carriage 17.5 days</td>
</tr>
<tr>
<td>Pt2</td>
<td>+</td>
<td>+</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>Gct carriage 10.5 days then censored</td>
</tr>
<tr>
<td>Pt3</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>Gct carriage 17.5 days</td>
</tr>
<tr>
<td>Pt4</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>Will be excluded from the analysis</td>
</tr>
</tbody>
</table>

Maximum gametocytaemia is defined as the maximum density of gametocytaemia recorded over the whole follow-up period. Only patients with all available counts will be included.

The area under the gametocyte density time curve (AUC) will be calculated using trapezoid rule and the actual gametocyte counts at days 0,7,14,21,28, and any later follow-up data. Patients with all counts available and patients with a missing count between two positive counts (the missing count, level of gametocytaemia will be interpolated from the two neighbouring values using simple regression) will be included in this analysis.

The normalised area under the gametocyte density time curve (AUC) will be used as a measure of gametocyte clearance. This will be calculated in the same way, but instead of gametocyte counts the % of maximum count will be used (i.e. gametocyte counts divided by the maximum gametocytaemia over study duration).
As a measure reflecting overall infectivity of an individual, a range of methods will be explored including the area under the curve (AUIC) of the probability of infecting a biting female anopheleline mosquito and time estimated and expressed in infectivity days, (Stepniewska et al. 2008, Bousema et al. 2012, Churcher et al. 2013).

Rate of gametocyte clearance will be defined in patients with gametocytaemia on enrollment as

1. a slope of the final decline in log gametocytaemia in patients with daily counts (provided there is enough data and the model is supported by the data)

2. gametocytes elimination rate of gametocyte as described by Bousema et al (T. Bousema et al. 2010)) in all patients with counts at d0,3,7,14.

7. Study and patient characteristics

The following baseline characteristics will be included in the analysis:

Site: Transmission intensity, degree of resistance etc

Patient: age, sex, weight, nutritional status, past history of malaria, history of fever, temperature on enrollment, anaemia

Drug: artemisinin derivative and its dose, partner drug and its dose, supervision of drug intake (full or partial), co-administration with fat and date of admission

Laboratory: baseline parasitaemia, species (Pf versus Pf mixed infection), haemoglobin,..

Anaemia will be defined according to WHO guidelines (WHO 2011). In studies with haematocrit measured instead of haemoglobin, haematocrit will be converted to haemoglobin using the following relationship:

\[ \text{Hematocrit (ht)} = 5.62 + 2.60 \times \text{Haemoglobin (hb)} \] (Lee et al. 2008)

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 (WHO 2006). 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 had been directly observed, partially supervised if at least the 3 morning doses had been observed, and not-supervised if fewer doses were observed.

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

WHO definitions of efficacy outcome will be used (WHO 2009).

For each patients early parasitological response will also be evaluated in the form of (a) parasite half life estimated by WWARN PCE tool (Flegg et al. 2011); (b) positivity on Day 2; (c) positivity on Day 3; (d) parasite half life estimated from daily counts, depending on the available data (WWARN Parasite Clearance Study Group, to be submitted).

For each study, study locations/sites will be recorded. Each location will be categorised into (a) low and high transmission settings based on the observed study site PCR confirmed reinfection rate, and the malaria endemicity estimates obtained for study sites and year from the Malaria Atlas Project (Gething et al. 2011) and (b) according to geographical region (Africa, Asia, and S. America).

To account for different methods of counting gametocytes, number of high power fields reviewed before diagnosis of negative made and the number of parasites seen on the thick film per specific number of white blood cells will be captured for each study included in the analysis.

Four levels of sensitivity for gametocyte detection will be defined as follows:

1. Slides were specifically read for gametocytes, reading ≥100 microscopic high power fields /≥1000 WBC (highest sensitivity)
2. Slide readers were specifically instructed to record gametocytes but slides were primarily read for asexual parasites; ≥100 microscopic high power fields/≥1000 WBC were read
3. Slide readers were specifically instructed to record gametocytes; 50-99 microscopic high power fields/500-999 WBC were read
4. Slide readers were not specifically instructed to record gametocytes or the number of examined high power fields <50 or the number of WBC was <500

8. Outline of Statistical Analysis

1. An overall summary and study profile of all the studies uploaded to WWARN repository will be presented and studies in which full follow-up gametocyte counts are available will be ascertained.
2. A summary of key study characteristics
3. Summary of patient baseline characteristics
4. Gametocytes carriage at 0, 7, 14, 21, 28, and any subsequent follow-up days by ACT treatments. These prevalences will be subgrouped for patients with (100% carriage on day 0) and without gametocytes on enrollment (0% carriage on day 0).

5. AUC for gametocyte carriage will be presented by ACT treatments, separately for patients with and without gametocytes on enrollment.

6. AUC for gametocyte carriage normalized by maximum gamatocytaemia will be presented by ACT, separately for patients with gametocytes on enrollment and patients who developed gametocytes after treatment.

7. Univariate and multivariate logistic regression for presence of gametocytes before initiation of treatment, with random effects for study site. All treatments will be included and the parameters outlined in section 7 assessed.

8. Univariate and multivariate logistic regression for appearance of gametocytes after treatment with ACT, with random effects for study site. Possible risk factors: patient baseline characteristics, transmission intensity, dose, ACT, sensitivity level, asexual parasite prevalence on day 3, parasite clearance time, recrudescence, new infection. Only patients with no gametocytes on enrolment will be included.

9. Duration of gametocyte carriage will be modelled as time to gametocyte clearance, provided enough variation will be observed in this endpoint. Univariate and multivariate models will be fitted, with random effects for study site, seperately for patients with gametocytes on enrolment and patients who developed gametocytmaemia after starting treatment. If majority of patients carry gametocytes 1 or 2 weeks, logistic regression will be performed. Possible risk factors: patient baseline characteristics, parasitaemia measures, transmission intensity, dose, ACT, sensitivity level, recrudescence, new infection.

10. Time to gametocytmaemia will be modelled as time to the first recorded positive gametocyte count in patients with no gametocytmaemia on enrollment and treated with ACT. Univariate and multivariate models will be fitted, with random effects for study site. Possible risk factors: patient baseline characteristics, parasitaemia measures, transmission intensity, dose, ACT, sensitivity level, recrudescence, new infection.

11. In patients with daily gametocytes counts, the kinetics of gametocyte clearance and association with kinetics of parasite clearance will be assessed.

12. Plot of gametocytes AUC (normalised and untransformed) versus parasite load/clearance measures : AUC (normalised and untransformed), PCE half life.

13. Plot of gametocyte clearance rate versus PCE half life.
9. Statistical Methodology

9.1. Descriptive statistics

Descriptive statistics will use mean and sd if data are normally distributed, geometric mean and range if data are log-normally distributed (as assessed by Shapiro-Wilk test), and as median and range otherwise.

9.2. Survival regression models for duration of gametocyte carriage and time to gametocytaemia

Duration of gametocyte carriage and time to gametocytaemia will be modelled using survival data regression.

For duration of gametocyte carriage, baseline will be defined as time of appearance of gametocytes. Time to clearance will be measured as time between baseline and first negative counts recorded. Patients who have the last available measurement before clearance of gametocytes or who have missing gametocyte counts after the positive counts will be treated as censored at the time of last recorded gametocytaemia.

Time to gametocytaemia will be measured from the enrolment until the first positive gametocyte count. Patients with missing intermittent measurements will be excluded.

For both models the same methodology will be applied. Random effects in the form of frailty parameters will be used to adjust for study-site effect (Glidden and Vittinghoff 2004). 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 (Collett 2003). In the Cox regression model, the proportional hazard assumption will be tested based on Schoenfeld residuals (Schoenfeld 1982). 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, as described in the section below.

9.3. Modelling gametocyte elimination rate constant

The following model (Bousema, 2010) will be fitted to gametocyte daily counts:

\[ G(t) = e^{-\mu t}G_0 + \rho S_0 (e^{-\mu t} - e^{-\rho t})/(\rho - \mu) \]
where $G_0$ is the density of circulation gametocytes on day 0, $S_0$ is the density of the sequestered gametocytes population on day 0, $\mu$ is the rate of elimination of gametocytes and $\rho$ is the rate of release of gametocytes from a sequestered population. Parameters $S_0$, $\rho$, $\mu$ will be estimated from the data together with their random effects for each patient, data permitted. Effect of the following covariates on these parameters will be evaluated: age, treatment, and transmission intensity.

9.4. **Model selection for risk factors**

For any regression model, the following strategy will be employed to determine independent risk factors. Initially all possible risk factors will be examined in the univariate model, and will be included in model building in the multivariate analysis. Model with known confounders will be fitted first (asexual parasite density, age in categories, transmission intensity baseline model). Variables and covariates will then be added to the baseline model in a stepwise forward fashion and the 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.

10. **Tools**

All statistical analyses will be carried out using R 2.14.0 released on 2011-10-31 by The R Foundation for Statistical Computing or Stata 13. However, when equivalent statistical methods are applied, changing the use of statistical software does not require amendment of this SAP.
11. References


## 12. Annex

### A.1 List of available covariates

<table>
<thead>
<tr>
<th>Description</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>WWARN Status for Pf Adj</td>
<td>Primary Response</td>
</tr>
<tr>
<td>WWARN Status for Pf UnAdj</td>
<td>Primary Response</td>
</tr>
<tr>
<td>Outcome (ETF, LTF etc)</td>
<td>Secondary Response</td>
</tr>
<tr>
<td>History of Fever (0/1) at inclusion</td>
<td>Baseline Variable</td>
</tr>
<tr>
<td>Severe Malaria at inclusion</td>
<td>Baseline Variable</td>
</tr>
<tr>
<td>Haemoglobin at inclusion</td>
<td>Baseline Variable</td>
</tr>
<tr>
<td>Falciparum density at Inclusion</td>
<td>Baseline Variable</td>
</tr>
<tr>
<td>Gamf (/µL) at inclusion</td>
<td>Baseline Variable</td>
</tr>
<tr>
<td>Max Temp Day0</td>
<td>Baseline Variable</td>
</tr>
<tr>
<td>D0 Ht&lt;20%</td>
<td>Baseline Variable</td>
</tr>
<tr>
<td>Age in Years</td>
<td>Baseline Variable</td>
</tr>
<tr>
<td>Gender</td>
<td>Baseline Variable</td>
</tr>
<tr>
<td>Weight</td>
<td>Baseline Variable</td>
</tr>
<tr>
<td>Antimalarial in last 28 days</td>
<td>Available Variable</td>
</tr>
<tr>
<td>Parasite density at Inclusion</td>
<td>Available Variable</td>
</tr>
<tr>
<td>Hb and/or Ht at any day</td>
<td>Available Variable</td>
</tr>
<tr>
<td>Temperature at any day</td>
<td>Available Variable</td>
</tr>
<tr>
<td>Asexual parasitaemia (counts and/or presence) at any day</td>
<td>Available Variable</td>
</tr>
<tr>
<td>Sexual parasitaemia (counts and/or presence) at any day</td>
<td>Available Variable</td>
</tr>
<tr>
<td>Max Falciparum Asexual parasitaemia on Day1</td>
<td>Available Variable</td>
</tr>
<tr>
<td>Max Falciparum Asexual parasitaemia on Day2</td>
<td>Available Variable</td>
</tr>
<tr>
<td>Max Falciparum Asexual parasitaemia on Day3</td>
<td>Available Variable</td>
</tr>
<tr>
<td>Max Temp Day1</td>
<td>Available Variable</td>
</tr>
<tr>
<td>Max Temp Day2</td>
<td>Available Variable</td>
</tr>
<tr>
<td>Max Temp Day3</td>
<td>Available Variable</td>
</tr>
<tr>
<td>Treatment information e.g. arm/tablets/supervision/fat etc</td>
<td>Available Variable</td>
</tr>
<tr>
<td>Dosing method (single day, broken down over days etc.)</td>
<td>Available Variable</td>
</tr>
<tr>
<td>Total mg/kg dose at each day of dosing regimen</td>
<td>Available Variable</td>
</tr>
<tr>
<td>Total mg/kg dose during course</td>
<td>Available Variable</td>
</tr>
</tbody>
</table>