

Statistical Properties of Parasite Density Estimators in Malaria

Imen Hammami^{1*}, Grégory Nuel¹, André Garcia²

1 Department of Applied Mathematics (MAP5), UMR CNRS 8145, Paris Descartes University, Paris, France, **2** Department of Mother and Child in Relation to Tropical Infections, UMR 216, Institut de Recherche pour le Développement, Paris, France

Abstract

Malaria is a global health problem responsible for nearly one million deaths every year around 85% of which concern children younger than five years old in Sub-Saharan Africa. In addition, around 300 million clinical cases are declared every year. The level of infection, expressed as parasite density, is classically defined as the number of asexual parasites relative to a microliter of blood. Microscopy of Giemsa-stained thick blood films is the gold standard for parasite enumeration. Parasite density estimation methods usually involve threshold values; either the number of white blood cells counted or the number of high power fields read. However, the statistical properties of parasite density estimators generated by these methods have largely been overlooked. Here, we studied the statistical properties (mean error, coefficient of variation, false negative rates) of parasite density estimators of commonly used threshold-based counting techniques depending on variable threshold values. We also assessed the influence of the thresholds on the cost-effectiveness of parasite density estimation methods. In addition, we gave more insights on the behavior of measurement errors according to varying threshold values, and on what should be the optimal threshold values that minimize this variability.

Citation: Hammami I, Nuel G, Garcia A (2013) Statistical Properties of Parasite Density Estimators in Malaria. PLoS ONE 8(3): e51987. doi:10.1371/journal.pone.0051987

Editor: Érika Martins Braga, Universidade Federal de Minas Gerais, Brazil

Received: January 16, 2012; **Accepted:** November 14, 2012; **Published:** March 14, 2013

Copyright: © 2013 Hammami et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors are funded by Centre National de la Recherche Scientifique (National Center for Scientific Research) (CNRS) PhD Grant. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: imen.hammami@parisdescartes.fr

Introduction

Malaria is caused by one or a combination of four species of *Plasmodia*: *Plasmodium falciparum*, *P. vivax*, *P. malariae*, and *P. ovale*. *P. falciparum* is the *Plasmodium* species responsible for 85% of the malaria cases and causes the most severe form of the disease. The three other species are less common and less dangerous. The usual and most reliable diagnosis of malaria is microscopic examination of blood films [1–6]. Two sorts of blood films can be used, thin or thick films, sampled from a finger prick. The thin smear is air-dried for 10 minutes. After drying, it is fixed in methanol and stained with Giemsa. Unlike the thin smear, the thick smear is dried for 30 minutes, but not fixed with methanol. It is de-hemoglobinized in water, and then stained with Giemsa [7]. Thin films allow species identification, because the parasite's appearance is better preserved. Thick blood smears are most useful for detecting the presence of parasites because they allow examining a larger sample of blood [8,9]. Thick films are hence more sensitive than thin films in case of low levels of infection and are therefore used to detect infection, and to estimate parasite density (PD) [1,10,11]. The level of infection, expressed as parasite density is classically defined as the number of asexual forms of parasite relative to a blood volume (e.g. microliter) or a percentage of white blood cells (WBCs).

Parasite density estimation methods usually involve threshold-based counting techniques [11–17]. Threshold definition and values may vary from one method to another. In some cases, parasites are counted in relation to the number of microscopic

high power fields (HPFs), defined as oil immersion microscopic fields ($\times 1,000$), and in other cases parasites are counted according to a fixed number of leukocytes (WBCs). In the first cases, the methodology is: if less than n parasites are counted in the m first HPFs, then do this, else do that. In the second ones, the m first HPFs are replaced by “when ℓ leukocytes are counted”, then do this, else do that. Conversion to counts per microliter then depends on the assumption that there is an average of 8,000 WBCs per microliter of blood.

Epidemiological interpretations must rely on solid evidence, hence the importance of reproducibility for parasite density data. However, all the methods used to determine PD potentially induce variability. To deal with this potential inaccurate estimation of PD, research teams tend to analyze more slides and subjects. By taking duplicate readings or larger sample sizes, we can statistically improve our knowledge of the PD being measured. Then, we can decrease the variability in microscope slide readings and improve the accuracy and reproducibility of the measurements. However, one of the problems the research teams have to deal with is that during large scale studies the number of thick blood smears performed can be greater than 10,000. Then, the repetition of the microscope slide examination leads to an important cost overrun in terms of both money and time. One may wonder whether such practices have a significant interest for the final results. With low parasitemias, it is probably worth the effort of reading more slides. But in some situations it is not needed, for example, with large parasitemias levels.

To our knowledge, none of the studies of variability have dealt with the sampling error generated by the threshold-based counting techniques or evaluated the impact of the existing threshold values in endpoint measurements. In addition, the accuracy and consistency of these methods have largely been overlooked. Furthermore, there is no general agreement on the optimal method for estimating parasite density according to threshold values. Further experimental evidence is needed to determine which parasite counting technique is most accurate, reproducible, and efficient. The aim of this paper was to explore the variability of four frequently used threshold-based counting methods of determination of PD. For each of these methods, we assessed the consequences that a modification of the threshold can have on variability.

Materials and Methods

Threshold-based counting techniques

PD estimates accuracy vary significantly depending on the methodology from which they are derived. The estimation method differs from one health care organisation to another. To summarize, there are four basic types of threshold-based counting techniques commonly used in epidemiological surveys.

Dowling & Shute (1966) [12] propose to count parasites in 200 consecutive HPFs in thick blood film [7,18]. They consider that the volume of blood corresponding to 200 HPF of a thick film is 0.4 μl , and that examining 200 HPFs is the best compromise between the need to reduce the risk of missing parasites and the need to minimize the reading time. This method will be referred to as Method A. Here, the number of HPFs read is the threshold value. We investigate the influence of this number on the reliability of the PD estimation.

Trape (1985) [11] proposes to count the total number of parasites per 200 WBCs. The average value of 8,000 WBCs per μl is accepted as reasonably accurate by *The World Health Organization* (WHO) [17]. Then, the number of parasites counted is multiplied by 40 to give the number of parasites per μl [9,19,20]. This method will be referred to as Method B. In this method, only one threshold value is specified, which is the number of leukocytes seen ℓ . We are interested in how the value of ℓ affects measures of variability.

According to the WHO recommendations [17], the number of parasites should be counted on one tally counter and the number of white blood cells should be counted on a second one. The number of parasites and white blood cells counted depend on how numerous the parasites are. The lower the number of parasites counted, the higher the number of WBCs that should be counted. Parasites are counted until 200 WBCs have been seen. If 100 or more parasites are found, the number of parasites per 200 WBCs is then recorded. Else, counting should be continued up to 500 WBCs. This method will be referred to as Method C. Three parameters are specified: the required number of parasites p , the required number of leukocytes in the first step ℓ_1 , and the required number of leukocytes in the second step ℓ_2 . Modeling, estimating and validating multidimensional distribution functions cast difficult problems, both conceptual and technical. For that reason, it is more convenient to fix $\ell_2=500$ and to study the method's performance by varying the two parameters p and ℓ_1 . Hence, we obtain the influence of adding an extra threshold value on the final estimation.

Finally, the last method presented in this paper was used during a research program conducted in the *Tori Bossito* area in *Southern Benin*. In this program, the PD is determined by simultaneously counting parasites and leukocytes. The counting stops when either

500 WBCs or 500 parasites are seen whichever comes first [15]. This method will be referred to as method D. Two parameters are specified: the required number of parasites p and the required number of leukocytes ℓ . We analyze the performance of the method with respect to effectiveness and efficiency for different values of parameters p and ℓ .

Unlike methods A and B, methods C and D are adaptive methods. In these methods, counting stops when parasites are found in sufficient number. Hence, their cost is reduced for high parasitemias.

Measures of variability

The source and scale of measurement error depends on several parameters, such as sample preparation, staining process, counting technique, microscopist performance, etc. However, variation of parasite density within a slide is expected even when prepared from a homogeneous sample [21]. The sampling variability is a source of interest when studying the efficiency of estimation methods. It refers to the different values which a given function of the data takes when it is computed for two or more samples drawn from the same population. In this paper we are interested in the sampling errors and biases induced by threshold-based counting techniques and more particularly in the impact of threshold values in endpoint measurements.

Let θ be the parameter that denotes the real value of the PD per microliter of blood and let $\hat{\theta}$ be its estimate. Since $\hat{\theta}$ is a random variable, it can never be said with certainty that this estimate is close to the true value of θ . For that reason, we consider its statistical properties, that is, its probability distribution $P(\hat{\theta})$, or some restricted aspects thereof. Here, we focus on variability measures: mean error (ME), coefficient of variation (CV) and false negative rate (FNR).

Mean error. In order to define the variability measures, we need to introduce the concept of mathematical expectation. The expected value of the estimator $\hat{\theta}$ denoted as $\mathbb{E}(\hat{\theta})$ is an average taken over all possible values of $\hat{\theta}$. Suppose $\hat{\theta}$ takes value s_1 with probability $p_1 = P(\hat{\theta}=s_1)$, value s_2 with probability $p_2 = P(\hat{\theta}=s_2)$, and so on, up to value s_n with probability $p_n = P(\hat{\theta}=s_n)$. Then the expectation of $\hat{\theta}$ is defined as

$$\mathbb{E}(\hat{\theta}) = \sum_k s_k p_k.$$

The sampling bias occurs when the true value (in the population) differs from the observed value (in the study) due to a flaw in the sample selection process. An estimator *bias* is the difference between the estimator's expected value $\mathbb{E}(\hat{\theta})$ and the true value of the estimated parameter θ . Hence, in computing the bias induced by different counting techniques, we used $\text{bias} = \sum_k s_k p_k - \theta$. An estimator with zero bias is called *unbiased*. Mean error is the bias expressed as a percentage of θ , i.e.

$ME = \frac{\text{bias}(\hat{\theta})}{\theta}$. It provides a measure of the magnitude of the bias and allows comparing different methods.

Coefficient of variation. A measure of the sampling error is the standard deviation which is the square root of its variance. Standard deviation is a measure of dispersion from the mean, or the expected value and it is commonly used to compute confidence intervals in statistical inferences. The reported margin of error is typically about twice the standard deviation (1.96), the radius of a 95 percent confidence interval. Sampling variability can also be

expressed relative to the estimate itself through the coefficient of variation (CV), which is defined as the ratio of the standard deviation σ to the true value θ . In computing the CV induced by different estimation methods, we used

$$CV = \frac{\sqrt{\sum_k (s_k - \mathbb{E}(\hat{\theta}))^2 p_k}}{\theta}.$$

Then, CV is expressed as a percentage of θ .

False negative rate. The last measure of variability we study is the false negative rate (FNR), also known as Type II error or β error, which is the error of failing to reject a false null hypothesis. The false negative rate indicates the probability of a counting method to estimate PD as null when it is not, i.e. $P(\hat{\theta} = 0 | \theta > 0)$.

Cost. In addition to the variability, we are also interested in the cost-effectiveness of each method. We define the method's cost as the number of HPFs that has to be read to reach the threshold value. Once it is reached, we stop the examination of the smear. Depending on the method being used, the cost is based on the number of parasites (or WBCs) required to stop the reading of the thick blood smear. Let T denote the required number of HPFs to stop the counting. We first compute $P(T = t)$. Then, we express the Method cost as $\mathbb{E}(T)$.

Methodology

The following methodology is used to compute the three measures of variability (ME, CV and FNR) and to assess cost-effectiveness of methods. Firstly, the exact distribution of $\hat{\theta}$ is computed through recursive formulas. Secondly, based on this probability density function, measures of variability are derived. Finally, cost-effectiveness is defined for each method as the required number of HPFs that has to be read until the threshold is reached. A C++ program is used to implement these recursive formulas. The calculations are performed under two assumptions :

A1. The distribution of the thickness of the smear, and hence of the parasites within the smear, is homogeneous.

A2. The distribution of the parasites in the HPFs is uniform, and thus can be modeled through a Poisson distribution [21–23].

Notations. Let X_i be a random variable that represents the number of parasites in the i -th HPF. Suppose that X_i are independent and identically distributed (i.i.d.).

Under an assumption of uniformity, the number of parasites per field can be modeled using Poisson distribution (assumption A2). If the expected number of parasites per HPF is λ_p , then $X_i \sim \mathcal{P}(\lambda_p)$. Thus $\mathbb{E}(X_i) = V(X_i) = \lambda_p$.

Let Y_i be a random variable that represents the number of leukocytes in the i -th HPF. Suppose that Y_i are independent and identically distributed (i.i.d.). Leukocytes are supposed evenly distributed over the thick smear. Therefore, the number of leukocytes per field can be modeled using the Poisson distribution. If the expected number of parasites per HPF is λ_ℓ , then $Y_i \sim \mathcal{P}(\lambda_\ell)$. Thus, $\mathbb{E}(Y_i) = V(Y_i) = \lambda_\ell$.

- Let ϕ denote the parasite density per WBC.
- Let S_t be the sum of parasites in t consecutive HPFs.

Then, $S_t = \sum_{i=1}^t X_i \sim \mathcal{P}(t\lambda_p)$.

- Let R_t be the sum of leukocytes in t consecutive HPFs.

Then, $R_t = \sum_{i=1}^t Y_i \sim \mathcal{P}(t\lambda_\ell)$.

- Let U_p be the minimum number of HPFs required to obtain p parasites.

U_p can be expressed in terms of X_i as follows

$$U_p = \arg \min_t \left\{ \sum_{i=1}^t X_i \geq p \right\}$$

Probability of U_p is given by

$$\begin{aligned} P(U_p = t) &= P(S_t \geq p, S_{t-1} < p) \\ &= \sum_{s=0}^{p-1} P(X_t \geq p - s) \cdot P(S_{t-1} = s) \end{aligned} \tag{1}$$

- Let V_ℓ be the minimum number of HPFs required to obtain ℓ leukocytes.

V_ℓ can be expressed in terms of Y_i as follows

$$V_\ell = \arg \min_t \left\{ \sum_{i=1}^t Y_i \geq \ell \right\}$$

The probability mass function of V_ℓ is given by

$$\begin{aligned} P(V_\ell = t) &= P(R_t \geq \ell, R_{t-1} < \ell) \\ &= \sum_{r=0}^{\ell-1} P(Y_t \geq \ell - r) \cdot P(R_{t-1} = r) \end{aligned} \tag{2}$$

PD estimation. For method A, natural estimator of θ is used and the exact formulas of ME, CV and FNR are given. However, the estimation of θ is not straightforward for the remaining methods (B, C, D). Hence, recurrence formulas are used to derive variability measures.

Let $\hat{\theta}_A$ be the estimator of θ for Method A. Let n be the number of HPF read. Since $X_i \sim \mathcal{P}(\lambda_p)$ and X_i are iid, we have $S_n \sim \mathcal{P}(n\lambda_p)$. The number of parasite per field λ_p is then estimated by $\hat{\lambda}_p$ where $\hat{\lambda}_p = \frac{S_n}{n}$. Assuming the average amount of blood in each field as $0.002 \mu l$ [12], the PD is estimated by $\hat{\theta}_A = \hat{\lambda}_p \times 500$. Since $\mathbb{E}[\hat{\theta}_A] = \theta$, $\hat{\theta}_A$ is unbiased. Thus, the ME is null.

In order to evaluate the efficiency of this estimator, the variance is to be compared against the Fisher Information $I(\theta)$. The variance of this unbiased estimator is bounded by the inverse of the $I(\theta)$; namely the Cramer-Rao Bound (CRB). We show that the variance of the proposed estimation technique reaches the Cramer-Rao lower bound. Hence, $\hat{\theta}_A$ is an efficient estimator of θ (the proof is given in Section 1 in File S1).

The coefficient of variation (CV) is defined as the ratio of the standard deviation σ to θ , which is equal to $\frac{1}{\sqrt{n\lambda_p}}$.

In practice, false negatives occur when diagnosing by mistake PD as null after reading n HPFs, i.e. $P(S_n = 0)$, which gives $FNR = e^{-n\lambda_p}$.

For Method B, ϕ is estimated by $\hat{\phi}_B = \frac{S_{V_\ell}}{R_{V_\ell}}$. Then, θ is estimated by $\hat{\theta}_B = \hat{\phi}_B \times 8,000$.

To derive statistical properties of the PD estimate, we first need to compute the probability of seeing k parasites (resp. r leukocytes) in V_ℓ HPFs. These probabilities can be expressed as follows

$$\begin{aligned} P(S_{V_\ell} = k) &= \sum_t P(S_t = k) \cdot P(V_\ell = t) \\ P(R_{V_\ell} = r) &= \sum_t P(R_t = r) \cdot P(V_\ell = t) \end{aligned} \tag{3}$$

S_t and R_t are Poisson-distributed and the probability of V_ℓ is computed according to Equation (2).

The probability density function of $\hat{\phi}_B$ is

$$P(\hat{\phi}_B = s) = \sum_{\substack{k \\ r = s}} P(S_{V_\ell} = k) \cdot P(R_{V_\ell} = r)$$

Let p be the required number of parasites for Method C. Let V_{ℓ_1} be the minimum number of HPF required to obtain ℓ_1 leukocytes and V_{ℓ_2} be the minimum number of HPF required to obtain ℓ_2 leukocytes. Let U_{ℓ_1, ℓ_2} be the minimum number of HPFs required to obtain p parasites. The probability mass function of U_{ℓ_1, ℓ_2} is as follows

$$P(U_{\ell_1, \ell_2} = t) = P(V_{\ell_1} = t) \cdot P(S_{V_{\ell_1}} \geq p) + P(V_{\ell_2} = t) \cdot P(S_{V_{\ell_1}} < p)$$

Probabilities of V_{ℓ_1} and V_{ℓ_2} are computed according to Equation (2).

Probability of $S_{V_{\ell_1}}$ is computed according to Equation (3).

Then, θ is estimated as follows $\hat{\theta}_C = \hat{\phi}_C \times 8,000$, where

$$\hat{\phi}_C = \frac{S_{U_{\ell_1, \ell_2}}}{R_{U_{\ell_1, \ell_2}}}$$

The probability density function of $\hat{\phi}_C$ is

$$P(\hat{\phi}_C = s) = \sum_{\substack{k \\ r = s}} P(S_{U_{\ell_1, \ell_2}} = k) \cdot P(R_{U_{\ell_1, \ell_2}} = r)$$

For Method D, $U_{\ell, p}$ denotes the minimum number of HPF required to obtain either ℓ leukocytes or p parasites, i.e. $U_{\ell, p} = \min(V_\ell, U_p)$. The probability mass function of $U_{\ell, p}$ is as follows

$$P(U_{\ell, p} = t) = P(V_\ell = t) \cdot P(U_p > t) + P(U_p = t) \cdot P(V_\ell > t)$$

Probabilities of U_p and V_ℓ are computed according to Equation (1) and Equation (2).

Then, θ is estimated by $\hat{\theta}_D = \hat{\phi}_D \times 8,000$, where $\hat{\phi}_D = \frac{S_{U_{\ell, p}}}{R_{U_{\ell, p}}}$.

The probability density function of $\hat{\phi}_D$ is

$$P(\hat{\phi}_D = s) = \sum_{\substack{k \\ r = s}} P(S_{U_{\ell, p}} = k) \cdot P(R_{U_{\ell, p}} = r)$$

Validation study. Simulations are used to study the accuracy of our mathematical models, and to validate the theoretical results derived from estimators' probability functions. For the purpose of simulations, we predefine a data-generating model of θ . Given λ_p and λ_ℓ , PD data are sequentially generated for each HPF. In that way, random samples of θ are generated under the Poisson assumption. Then, we investigate properties of sample means, variances and FNR. We use the statistical software package R to perform 10,000 simulations. In each simulation step, we generate 1,000 random drawings of θ and we save the sample ME, CV, FNR and cost in a vector. In that way, we were able to investigate the results of all simulation steps. We compare the simulated results to the theoretical ones. Some results of our validation study are given in Section 2 in File S1. Simulations are computationally expensive. Then, it is burdensome to have to perform 10,000 simulations to estimate each PD value according to methods A, B, C and D. Hence, computing the exact distribution of θ is a most useful alternative. Codes used for generating the data will be provided to the reader upon request.

Colormaps. The recursive formulas described above are used to compute the exact distribution of variables. Each computation takes as input: λ_ℓ the number of leukocytes per field, λ_p the number of parasites per field and the threshold values (number of HPFs, number of WBCs, number of parasites). The outputs are ME, CV, FNR and cost values. This approach is computationally expensive due to recursive formulas that precisely compute probability of getting r WBCs (resp. k parasites) according to each counting technique. These probabilities are used to compute $P(\hat{\theta})$. Statistical properties of PD estimators are then derived. These data sets are gridded into colormaps where the values taken by a variable (ME, CV, FNR, cost) in a two-dimensional table (X,Y) are represented as colors. Each rectangle in this grid is a pixel (or a color sample). This program sets each pixel to a color index according to its coordinates. Each pixel has an X and Y position where the X coordinate is the parasite density value and the Y coordinate is the threshold value. The X axis spans the range of 0 to 20,000 parasite per μl (400 values). The Y axis ranges from 0 to 500 (500 values). Hence, we used a resolution of 400×500 pixels. Contour lines are overlaid over the colormaps. A contour line connects points where the function has constant value. Linear interpolation is used in generating contour data. A higher resolution is needed to achieve a smoother mapping and to avoid artifacts (jagged contours), which arise due to interpolation.

We believe that if data mappings are addressed simultaneously in a single framework, the resulting approach will facilitate visual comparisons of methods. At that point, we consider the problem of scale. The methods use either different numbers of arguments or different types of arguments. We got rid of this by expressing variability measures for all methods as functions of parasity density and WBCs count. To do so, we convert threshold values used in each method into WBCs count. For Method A, we assume an average of 8,000 WBCs per microliter of blood [17] and an average of 0.002 microliter of blood in each field [12]. The number of HPFs read n is then multiplied by $\lambda_\ell = 16$ WBCs to give the number of WBCs counted in n HPFs. For Method B, the threshold value is the number of WBCs counted ℓ . For Method C, we consider the case where $\ell_1 = \frac{p}{2}$ and we fix $\ell_2 = 500$. For Method D, we consider equal numbers of parasites p and leukocytes ℓ that have to be seen to stop the counting, hence $\ell = p$. These decisions are based on common use and can be considered reasonable assumptions. Moreover, this two-dimensional representation has the conceptual advantage of reducing the

number of arguments and ensures a common approach for assessing methods performance.

Results

In the following, parasitemia was categorized as either low ($PD \leq 100$ parasites/ μl), intermediate ($100 < PD < 10,000$ parasites/ μl) or high ($PD \geq 10,000$ parasites/ μl).

Method C involves three threshold values, which leads to a multidimensional problem. For this reason, we chose to express the parasite count in the first step as the half of the leukocyte count. We fixed the leukocyte count in the second step to 500 WBCs.

Impact of thresholds on variability measures

Mean error. The mean value of the Method A estimator equals the true value of the PD. Therefore this estimator is called an unbiased estimator. In addition to having the lowest variance among unbiased estimators (so called Minimum Variance Unbiased Estimator), this estimator also satisfies the Cramér-Rao bound, which is an absolute lower bound on variance for statistics of a variable and thus is an unbiased efficient estimator. Analytic proofs are given in Section 1 in File S1.

As shown in Figure 1, the ME of Method B only depends on threshold values. In Method B, parasites are counted until a fixed number of WBCs are seen and the number of parasites seen is not involved in the stopping rule of this counting process. Hence, the ME is independent from the PD. The colormap of the ME shows that the mean error decreases as threshold values increase. For instance, counting parasites until 400 WBCs instead of 200 WBCs decreases the bias by 0.5% of the parasite density. This can help choose the threshold value that allows to decrease the bias to a reasonable value.

For Method C, three parts can be distinguished. For $PD \leq 4,000$ parasites/ μl , contour lines are increasing functions of the PD and the thresholds. The darkest part on the map represents a constant ME. Due to the limited number of parasites in this area, counting is carried out until 500 WBCs are seen. Hence, Method C has similar behavior to Method B for $WBCs = 500$. By counting up to 500 WBCs, the mean error is fixed to 0.4%. For $1,000 < PD < 10,000$ parasites/ μl , ME values are represented by a set of bell-shaped density curves with a peak reached at $PD = 4,000$ parasites/ μl . 4,000 is half the standard number of WBCs per μl . For Method C, the number of parasites counted is half the number of leukocytes. For $PD \leq 4,000$ parasites/ μl , increasing the parasite density increases the leukocyte count for a fixed ME value. If the microscopist wants to estimate for constant ME a higher parasite density, he needs to count more leukocytes. A higher threshold value is then required due to the small number of parasites present in this area. For $PD > 4,000$, lower leukocyte counts are needed to maintain a constant ME value. A steady state will be reached afterwards whereby the ME is density independent. Due to the abundance of parasites, the ME only depends on the WBCs count. This steady-state region starts at $PD = 6,650$ parasites/ μl for $ME = 0.5\%$.

Note that the same ME level may be reached by more than one threshold value (eg. two contour lines for $ME = 10\%$).

For $PD \leq 6,000$ parasites/ μl , the ME generated by Method D is density independent (see Figure 1). In this interval, leukocytes are more abundant than parasites. Hence, parasites are counted until a predetermined number of leukocytes is reached. We notice that increasing the leukocyte count will not significantly reduce the bias. For instance, counting parasites until 400 WBCs are seen, instead of 200, decreases the bias only by 0.5% of the parasite

density. For $6,000 < PD < 10,000$ parasites/ μl , contour lines reach their minimum at $PD = 8,000$ parasites/ μl . In this area, the number of leukocytes per field (λ_L) and the number of parasites per field (λ_P) are very close. Lower threshold values are needed to maintain a constant ME. For high parasitemia, parasites are more numerous than leukocytes. The ME is therefore density dependent. If the microscopist wants to estimate for constant ME a higher parasite density, he needs to count more parasites.

Coefficient of variation. The CV of Method A is the inverse square root of λ_P times the number of fields (see Materials & Methods). If the counting does not exceed 12 HPFs (i.e. $WBCs \leq 200$), CV values are higher than 9.94% of the real PD for low and intermediate parasitemias (see Figure 2). Above 20,000 parasites/ μl , CV values are less than 3%. Counting up to 31 HPFs (i.e. $WBCs \approx 500$) instead of 12 HPFs (i.e. $WBCs \approx 200$), decreases the CV by approximately 10% of the parasite density.

For Method B, CV values lie midway between 10% and 20% of the PD for high parasitemias when $WBCs \geq 100$ (see Figure 2). Notice that increasing the number of WBCs counted will not significantly decrease the CV for high parasitemias.

For Method C, the vertical lines indicate that the CV only depends on PD for low parasitemias. Due to the small number of parasites, CV levels are obtained by counting parasites until 500 WBCs are seen. Notice that CV values exactly match those obtained by Method B with $WBCs = 500$. Figure 2 shows bell-shaped patterns for higher densities (for $1,000 < PD < 10,000$ parasites/ μl) with a peak reached at $PD = 4,000$ parasites/ μl . Along the same line as Method B, a constant CV level may be reached by more than one threshold value for $PD \geq 13,500$ parasites/ μl .

For Method D, the negative slope of the contour lines captures the indirect relationship between the threshold and the densities for $PD < 8,000$ parasites/ μl . For a fixed CV level, threshold values decline with density. In this interval, the counting stops when the fixed number of leukocytes (i.e. the threshold value) is obtained. The minimum is reached at $PD = 8,000$ parasites/ μl . For $PD > 8,000$ parasites/ μl , positively sloped CV curves reflect the direct relationship between the threshold and the PD. In this area, parasites are more abundant than leukocytes. Therefore, the counting stops when the fixed number of parasites (i.e. the threshold value) is reached. If the microscopist wants to estimate with the same level of precision (i.e. for constant CV) a higher parasite density, he needs to count more parasites. A higher threshold value is then required.

False negative rates. For Method A, FNR decreases exponentially with increasing number of fields (n) and increasing number of parasites per field (λ_P) (see Materials & Methods). If the counting does not exceed 20 HPFs (i.e. $WBCs \leq 320$), the probability of misdiagnosis is high for low parasitemia levels (see Figure 3). For intermediate densities, this probability is less than 1% when the threshold is above 30 HPFs. For high parasitemias, false negatives occur much less frequently ($< 0.001\%$). Despite unbiasedness and efficiency, this estimator generates a high number of false negatives when the problem is difficult (low parasitemia).

Figure 3 shows that the FNRs of Method B vary from 5% to 80% for low parasitemia levels. For intermediate densities, this probability is less than 5%. False negatives do not occur for high parasitemia levels.

For Method C, the FNRs are threshold independent for $PD \leq 200$ parasites/ μl and $WBCs > 20$. The number of false negatives arises from counting up to 500 WBCs. For $200 < PD \leq 2,000$ parasites/ μl and $10 \leq WBCs \leq 20$, FNR values varies from 0.001% to 0.5%. For $PD \leq 2,000$ parasites/ μl and

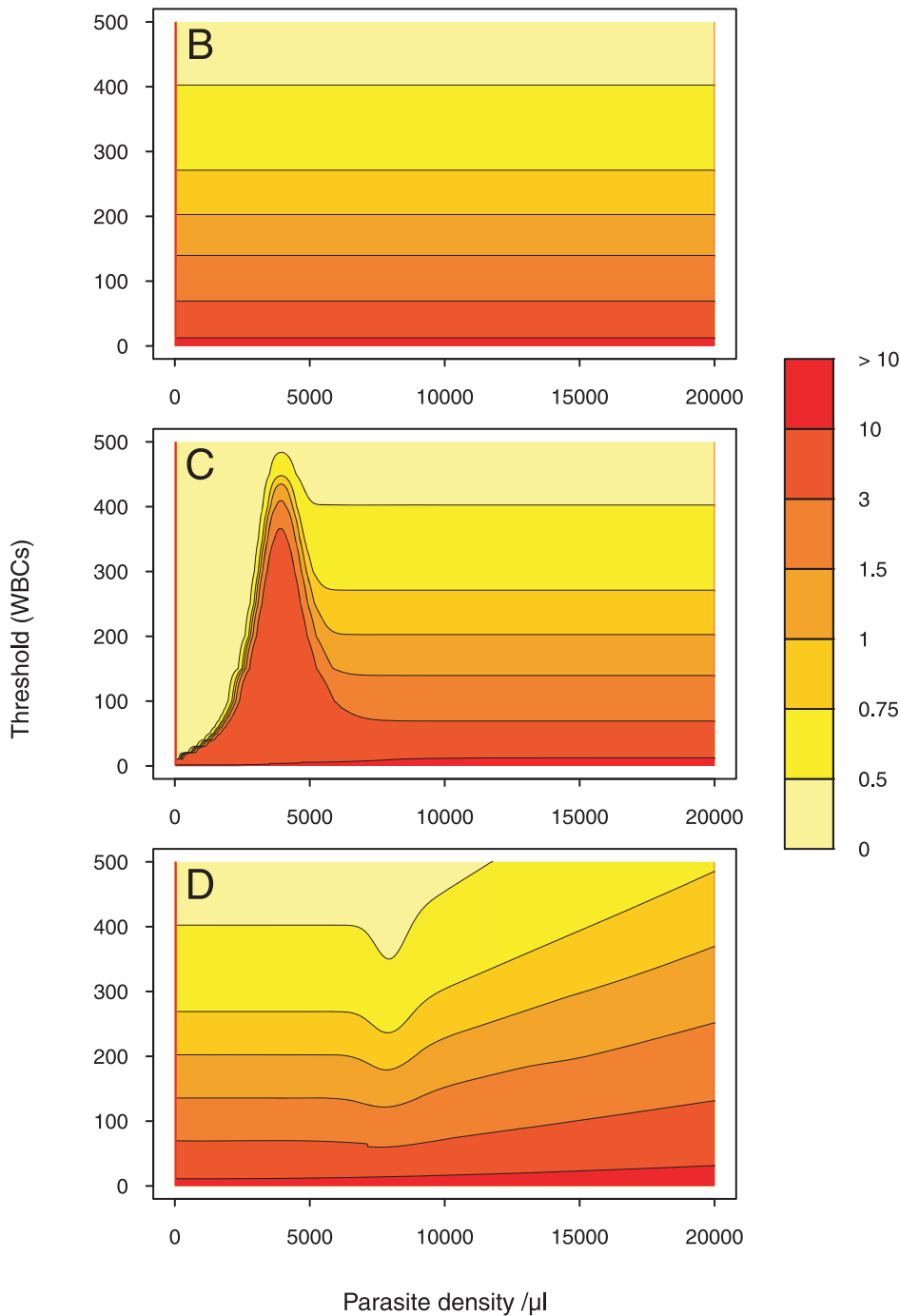


Figure 1. Mean error colormap. The colormap is drawn given a two-dimensional array of ME values. To allow for direct point-to-point numerical and visual comparison, we express the ME as a function of the parasite density (on the *x*-axis) and the WBC count (on the *y*-axis) in each of the four methods. Parasite density values are generated starting with 0, at increments of 50, and ending with 20,000. Threshold values (WBCs) are generated starting with 0, at increments of 1, and ending with 500. Then, each pixel is assigned a value that represents the ME-level. A color scale grading was applied to show levels. 7 degree intervals are depicted using a red-to-yellow colorspace with increasing intensity. We contour the ME at 0.5, 0.75, 1, 1.5, 3 and 10. The gaps between each pair of neighboring contour lines is filled with a color.
doi:10.1371/journal.pone.0051987.g001

WBCs < 10, FNR values are higher than 0.5%. False negatives do not occur for high parasitemia levels ($PD \geq 10,000$ parasites/ μ l).

For low parasitemias, we point out striking similarities between the FNRs in Method D and the FNRs in Method B. Due to the scarcity of parasites in this area, estimates are based on the leukocyte count in Method D.

Cost-effectiveness. Method A does not adapt to the variation of PD from one individual to another and costs a fixed HPFs number for all PD values.

The cost of Method B is an increasing linear function of the threshold values (see Figure 4). The cost here is independent from PD. This can be explained by the homogeneous distribution of

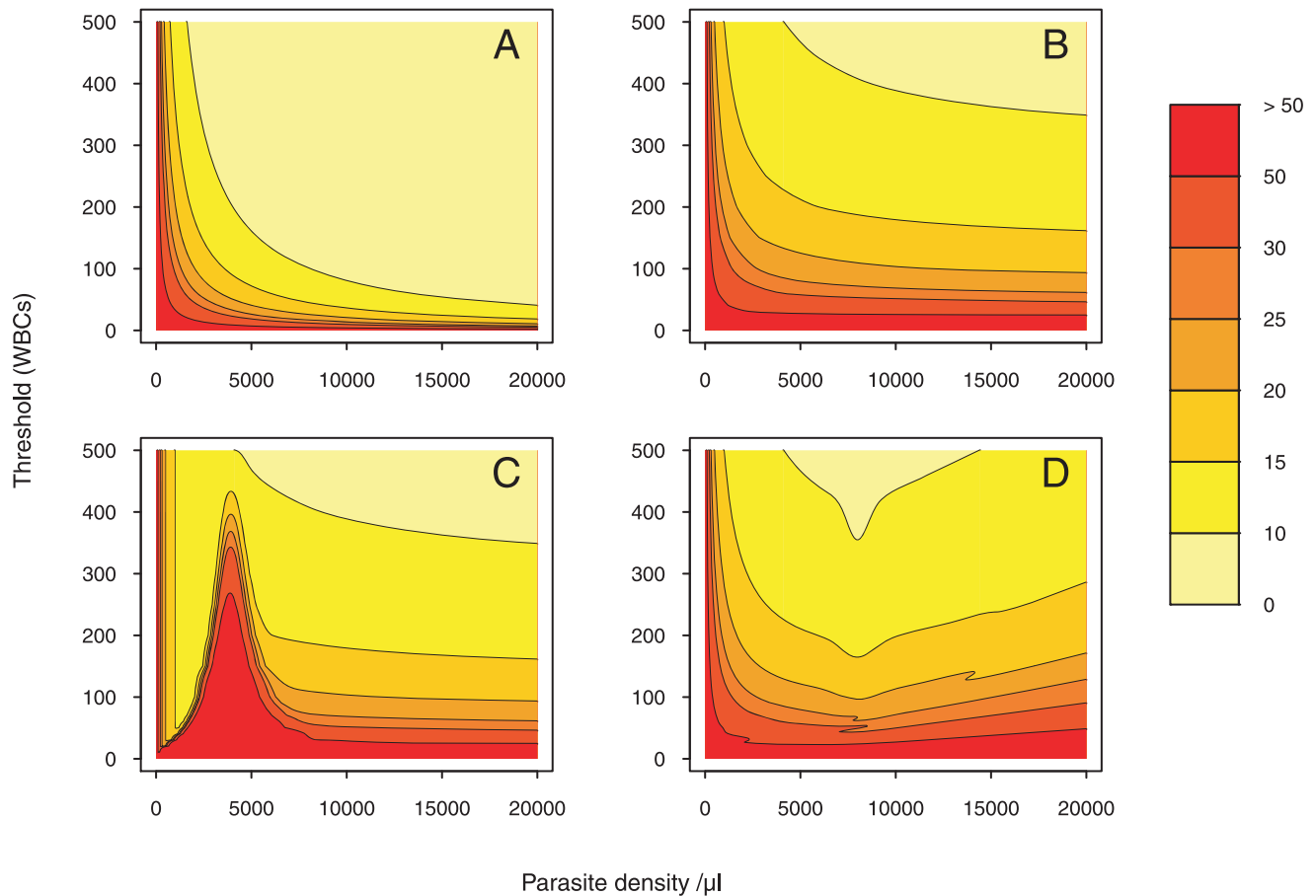


Figure 2. Coefficient of variation colormap. The colormap is drawn given a two-dimensional array of CV values. To allow for direct point-to-point numerical and visual comparison, we express the CV as a function of the parasite density (on the x -axis) and the WBC count (on the y -axis) in each of the four methods. Parasite density values are generated starting with 0, at increments of 50, and ending with 20,000. Threshold values (WBCs) are generated starting with 0, at increments of 1, and ending with 500. Then, each pixel is assigned a value that represents the CV-level. A color scale grading was applied to show levels. 7 degree intervals are depicted using a red-to-yellow colorspace with increasing intensity. We contour the CV at 10, 15, 20, 25, 30 and 50. The gaps between each pair of neighboring contour lines is filled with a color. doi:10.1371/journal.pone.0051987.g002

leukocytes within the fields. Since we assumed a fixed number of leukocytes per field (λ_ℓ), the number of fields needed will indeed be independent of the PD.

For low parasitemia levels and WBCs > 5 , the darkest color in the cost colormap of Method C indicates a constant cost of approximately 31 HPFs, which corresponds to the number of fields needed to reach 500 WBCs. For intermediate parasitemia levels, the cost varies depending on both the threshold value and how numerous the parasites are. The cost is independent from PD for high parasitemia levels. The number of fields needed is the ratio of WBCs to λ_ℓ .

Method D is highly adapted to parasitemia levels in terms of cost. For $PD \leq 8,000$ parasites/ μl , the cost is density independent. For $PD > 8,000$ parasites/ μl , the cost decreases with density for a fixed threshold value. In this interval, parasites are more abundant than leukocytes. A lower number of fields is then needed to reach a predetermined threshold value.

Methods comparison for three parasitemia levels

To explore similarities and differences in method behaviors, we look more closely at the statistical properties of PD estimates. We choose three cut-offs for low (100 parasites/ μl), intermediate (1,000 parasites/ μl) and high (10,000 parasites/ μl) parasitemias.

As Method A was shown to be unbiased, it was excluded from the ME analysis. As shown in Figure 5, Method B and Method D seem to have similar behaviors in terms of ME for low and intermediate parasitemias insofar as the two estimates are based on the leukocyte count in this density interval. For high parasitemias, Method B and Method C give the same results. The parasite count in Method C does not influence the accuracy of the method as long as parasites are numerous. For this reason, the two methods basically behave the same way.

To understand how the threshold values influence the variability of PD estimates, we plotted the CV according to threshold values. Figure 5 shows that the CV is highly sensitive to any variation of low thresholds (≤ 100). However, we see very few variations of the CV as threshold values increase (> 100). Both Methods B and D generate very close CV values for low and intermediate parasitemia levels. This result is expected since the number of WBCs seen is greater than the parasite number in the considered PD intervals. Hence, the two methods have the same stopping rules. For high parasitemias, Method B and Method C generate similar variability whereas Method A is significantly more precise than the other methods (B, C, D). However, Method A generates higher FNR for intermediate parasitemia than other methods when the count does not exceed 20 HPFs. For high PD

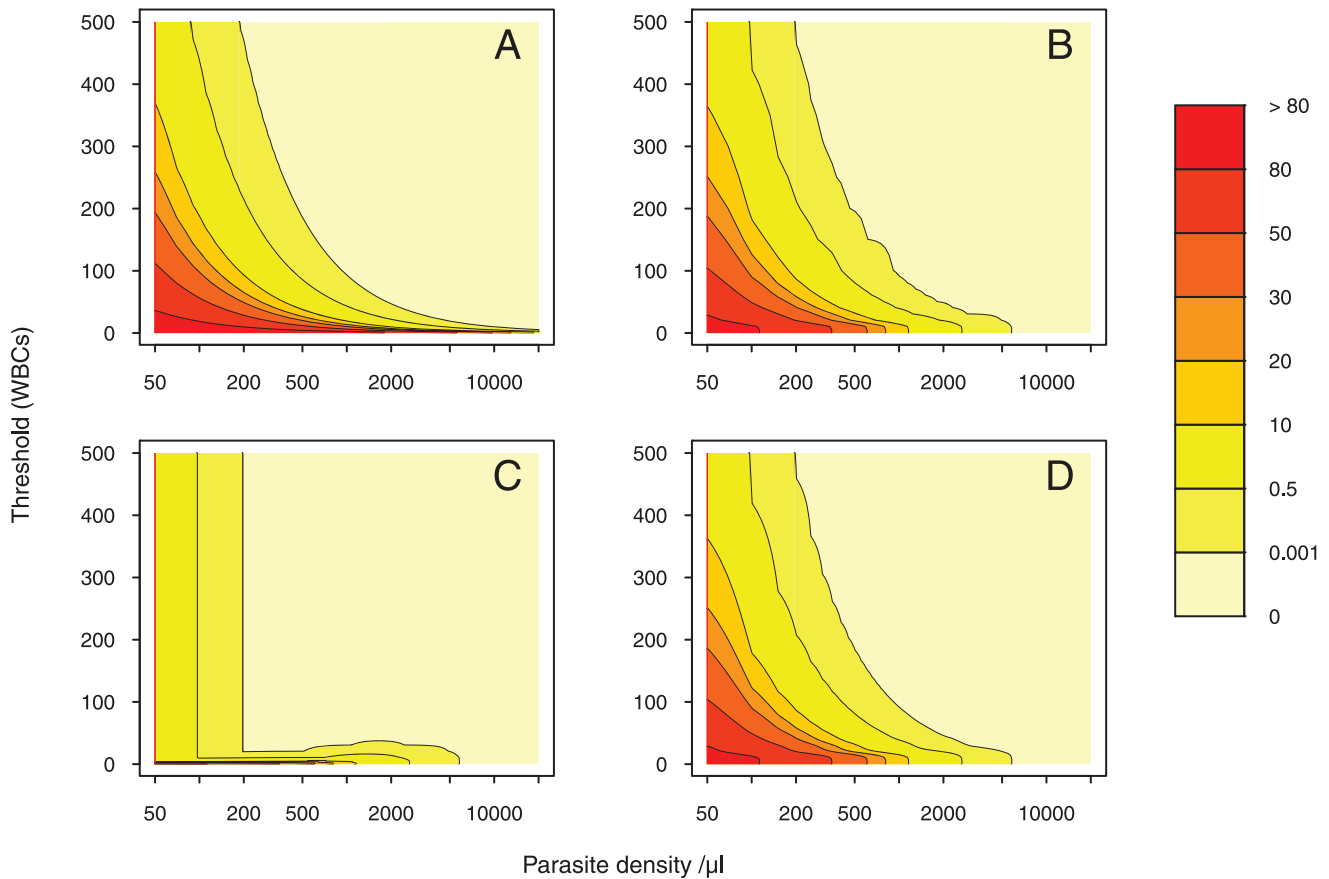


Figure 3. False negative rates colormap. The colormap is drawn given a two-dimensional array of FNR values. To allow for direct point-to-point numerical and visual comparison, we express the FNR as a function of the parasite density (on the *x*-axis) and the WBC count (on the *y*-axis) in each of the four methods. Parasite density values are generated starting with 0, at increments of 50, and ending with 20,000. Threshold values (WBCs) are generated starting with 0, at increments of 1, and ending with 500. Then, each pixel is assigned a value that represents the FNR-level. A color scale grading was applied to show levels. 8 degree intervals are depicted using a red-to-yellow colorspace with increasing intensity. We contour the CV at 0.001, 0.5, 10, 20, 30, 50 and 80. The gaps between each pair of neighboring contour lines is filled with a color. A logarithmic scale is used on the *x*-axis and a linear scale is used on the *y*-axis.
doi:10.1371/journal.pone.0051987.g003

levels, false negatives do not occur when the count exceeds 5 HPFs.

Figure 5 point out the high level of accuracy and precision performance of Method C for low and intermediate parasitemias. Thus, adding a supplementary stopping rule to the counting process and taking into account the parasite counts have enhanced the method performance, which raises questions regarding the repercussions in terms of cost-effectiveness. As shown in Figure 5, Method C is more expensive and time-consuming for low and intermediate parasitemias and requires constant cost (31.5 HPFs). As leukocytes are more present than parasites in the considered PD intervals the counting will be carried out until 500 WBCs are seen. Method A and B costs are density independent and increase linearly with threshold values. Method D outperformed the three other methods in terms of cost for high parasitemia levels.

Variability of measurements at equal cost-effectiveness

The duality between variability and cost illustrated in the previous section prompted a more detailed analysis of method performance differences at equal cost-effectiveness. In order to do this, we represent the variability measures as a function of cost. As shown in Figure 6, Methods B and D behave the same in terms of ME and CV for low and intermediate PD levels. For Method C,

ME, CV and FNR are density independent for low and intermediate parasitemias. For high PD levels, Methods B, C and D present similar results for ME, CV and FNR. Method A has the lowest CV values but generates higher FNR.

Methods comparison for standards threshold values

To identify both similarities and differences between the commonly used threshold-based counting techniques, we estimate ME, CV, FNR and cost as a percentage of PD according to three parasitemia levels (low, intermediate, high) for commonly used threshold values. We used 200 HPFs for Method A, 200 WBCs for Method B, 100 parasites and 200 WBCs for Method C and 500 WBCs or 500 parasites for Method D.

Table 1 shows that Method A is the most efficient method in terms of accuracy (ME) and precision (CV), but has an important cost (200 HPF). Conversely, Method B is less accurate and precise than Method A while much more cost-effective (12.5 HPFs). Method C and Method D present similar properties for low and intermediate parasitemia levels. In fact, Method C behaves as Method D with a fixed leukocyte count (500 WBCs) due to the scanty presence of parasites. Hence, the mean error is density independent in these PD intervals. For high PD levels, Methods B and C behave the same. Due to the abundance of parasites, the

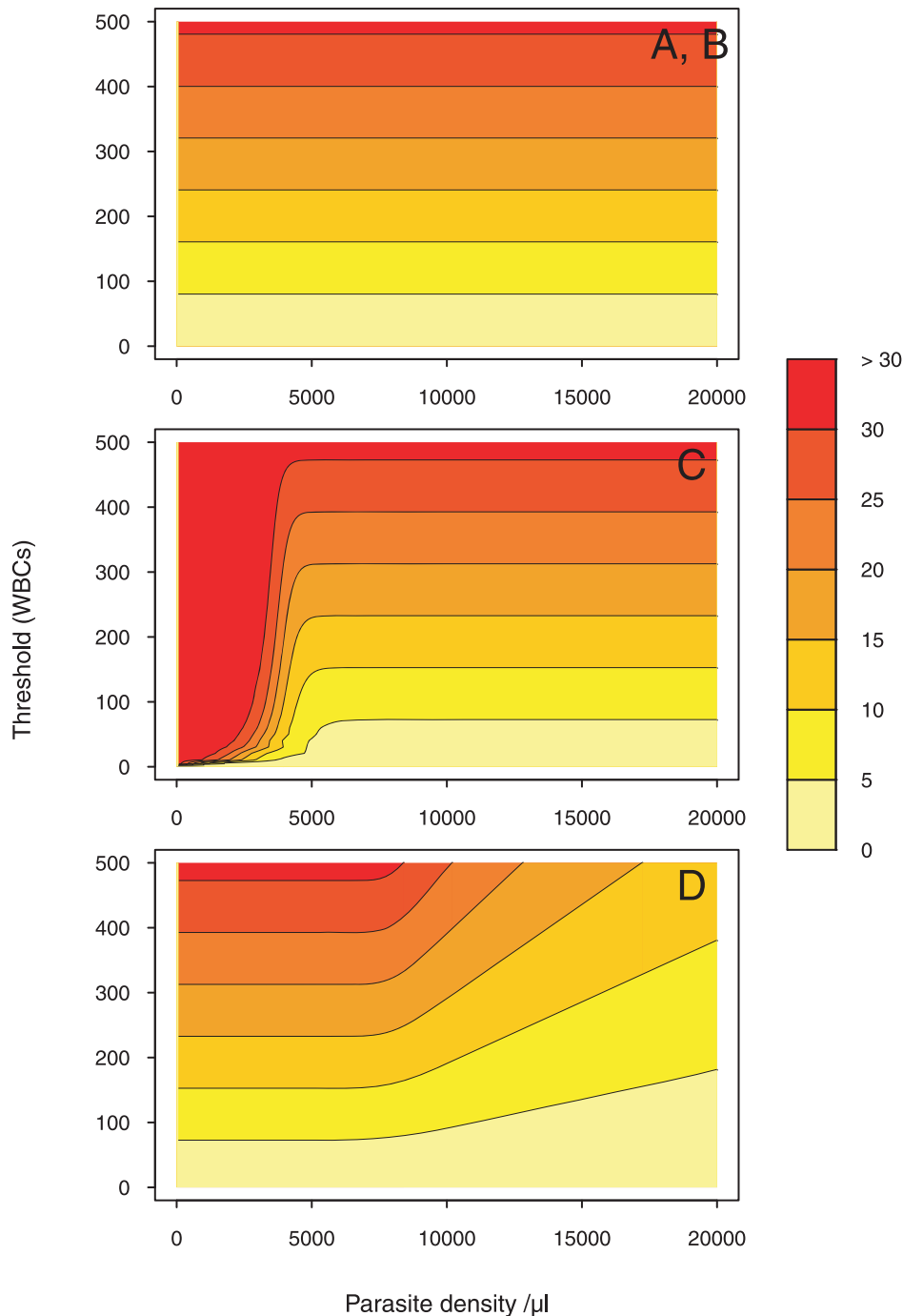


Figure 4. Cost-effectiveness colormap. The colormap is drawn given a two-dimensional array of cost values. To allow for direct point-to-point numerical and visual comparison, we express the cost as a function of the parasite density (on the x -axis) and the WBC count (on the y -axis) in each of the four methods. Parasite density values are generated starting with 0, at increments of 50, and ending with 20,000. Threshold values (WBCs) are generated starting with 0, at increments of 1, and ending with 500. Then, each pixel is assigned a value that represents the cost-level. A color scale grading was applied to show levels. 7 degree intervals are depicted using a red-to-yellow colorspace with increasing intensity. We contour the cost at 5, 10, 15, 20, 25 and 30. The gaps between each pair of neighboring contour lines is filled with a color. doi:10.1371/journal.pone.0051987.g004

enumeration is stopped when 200 WBCs are seen in both counting procedures. However, Method D is better suited to high parasitemia levels in terms of accuracy and precision compared to Methods B and C but results in up to a 2-fold increase in costs.

Discussion

To the best of our knowledge, this is the first study of threshold-based counting technique performance using the theoretical properties of PD estimators. We considered four commonly used threshold-based counting techniques, and assessed the perfor-

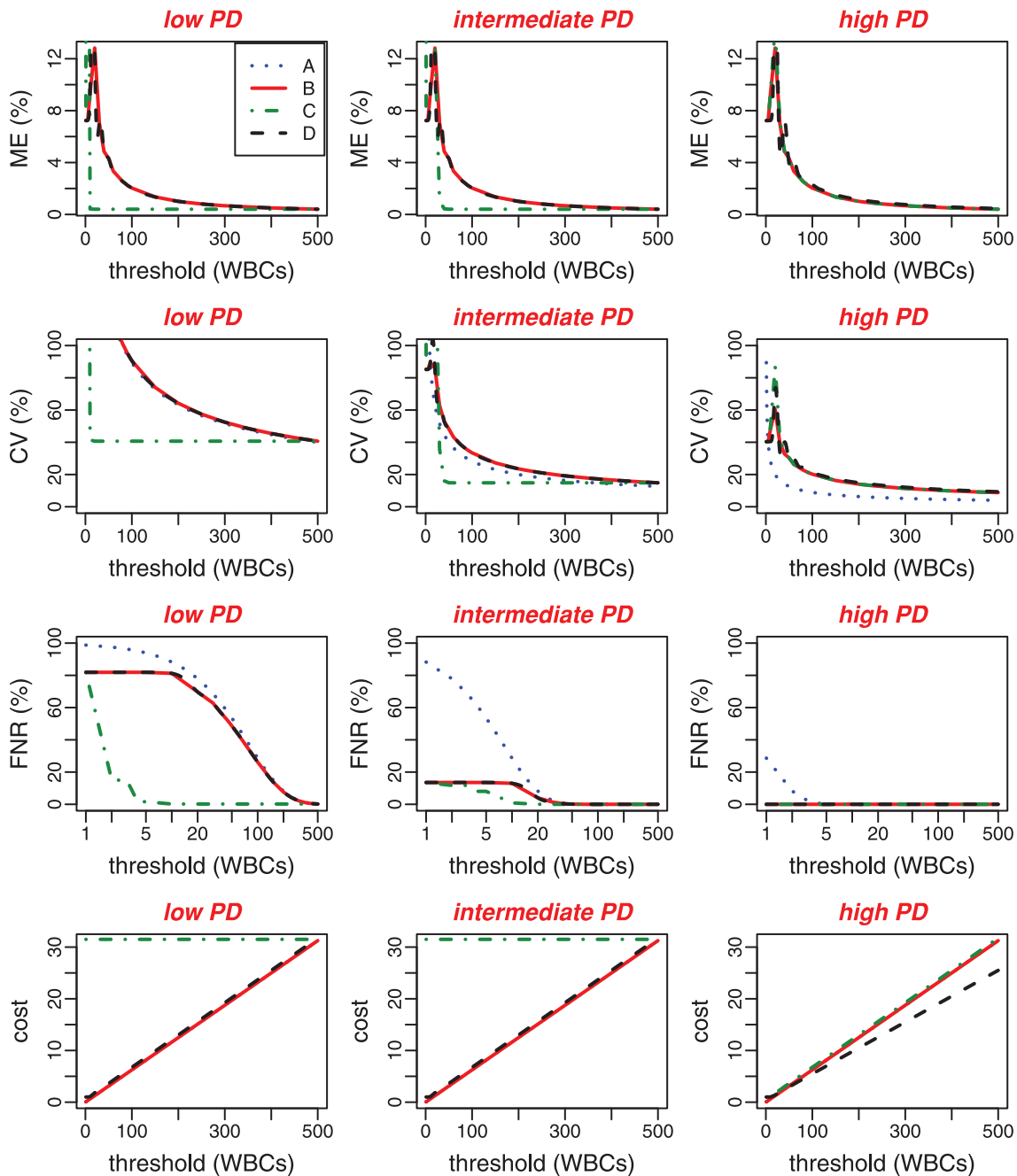


Figure 5. Statistical properties of PD estimators cut-offs according to threshold values for three PD levels : low (100 parasites/ μ l), intermediate (1,000 parasites/ μ l) and high (10,000 parasites/ μ l). Variability measures (ME, CV, FNR) and cost are expressed as functions of the WBCs count (threshold) for the four methods (A, B, C, D). This graph gives the required number of WBCs for each method according to an expected amount of variability or cost, and favours a direct comparison between methods in terms of WBCs count. A logarithmic scale is used on the x -axis for FNR.
doi:10.1371/journal.pone.0051987.g005

mances of these methods according to threshold values. These thresholds may be fixed or variable. We showed that adaptative methods are more efficient than the ones involving fixed threshold values. To define the theoretical properties of the estimators we hypothesized that the distribution of parasites within HPFs follows a Poisson distribution. We demonstrated that Method A estimator is unbiased and efficient. However, this estimator generates a high number of false negatives, especially for low parasitemia levels when the counting does not exceed few HPFs. Moreover, Method

A is time-consuming. We showed that the ME of Method B is independent from PD, and only depends on the threshold value. This helps to handle the amount of bias with an appropriate choice of the WBC threshold value. We showed that adding a new parameter to the stopping rules (the number of parasites seen) implies more accuracy and precision without increasing the method's cost for low and intermediate parasitemias. Method B and Method D have similar behaviors for low and intermediate parasitemia levels while Method D is more accurate and precise in

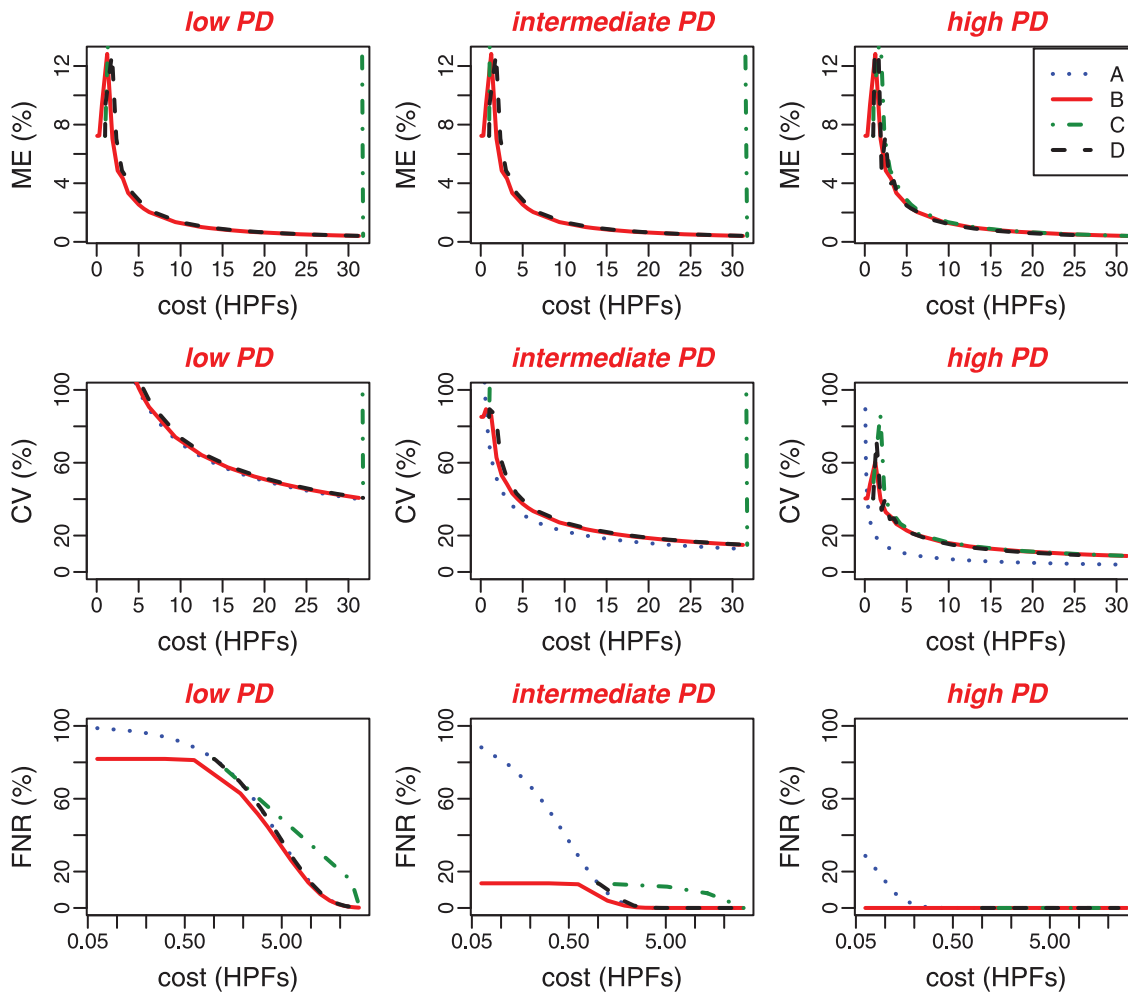


Figure 6. Statistical properties of PD estimators cut-offs according to methods cost for three PD levels : low (100 parasites/μl), intermediate (1,000 parasites/μl) and high (10,000 parasites/μl). Variability measures (ME, CV, FNR) are expressed as functions of the cost (the number of HPFs needed to stop the counting) for the four methods (A, B, C, D). This graph gives the cost for each method according to an expected amount of variability, and favours a direct comparison between methods in terms of cost. A logarithmic scale is used on the *x*-axis for FNR. doi:10.1371/journal.pone.0051987.g006

the considered PD intervals. For high parasitemia levels, Method B and Method C have similar behaviors and are more accurate and precise than Method D. However, for high parasitemias, Method D outperformed the three other methods in terms of cost. For each method, different threshold values may be fixed, which raises questions regarding the accuracy and reproducibility of these parasite counting techniques.

The importance of parasite density data reproducibility stems from the need for epidemiological interpretations to be based on solid evidence. However, variation of parasite density within a slide is expected even when prepared from a homogeneous sample [21]. The source and scale of measurement error (sample preparation, staining process, counting technique, microscopist performance) have been investigated. The notion of inter-rater reliability is a source of concern in this context. It refers to a metric for raters' consistency that measures the degree of agreement among raters. Many techniques were developed to measure inter-rater reliability. Some reports deal with the variability in the methods for detecting and counting parasites in thick smears. They attempt to evaluate the inter-rater reliability of malaria microscopy in epidemiological studies by looking at the variation of results due to the microscopist's reading. The variability of these

methods has been assessed using statistical approaches [11,12,20,21,24–29]. These methods used several criteria to assess the inter-rater reliability and to quantify the degree of agreement between malaria slide density readings. For continuous data, Analysis of Variance (ANOVA) is the method of choice. Bland & Altman (1986) [24] plotted the differences in log-transformed data versus average in mean counts. They expanded on this idea by plotting the difference of each point, the mean difference, and the confidence limits on the vertical axis against the average of the two ratings on the horizontal axis. The resulting Bland & Altman plot [24] demonstrates not only the overall degree of agreement, but also whether the agreement is related to the underlying value of the item. For instance, two raters might nearly agree in estimating the size of small items, but disagree about larger ones. Alexander et al. (2010) [21] assessed agreement between replicate slide readings of malaria parasite density using as criterion the repeatability, that is to say the value below which the absolute difference between results may be expected to lie with a 95% probability [30]. This metric is linked to Bland & Altman limits of agreement [24]. It is half the distance between the upper and lower limits of agreements. For nominal data, the kappa coefficient

Table 1. Threshold-based counting techniques comparison for low (100 parasites/ μ l), intermediate (1,000 parasites/ μ l) and high (10,000 parasites/ μ l) parasitemias.

Low parasitemia				
$\theta = 100$ parasites/ μ l				
Method	A	B	C	D
ME ($\%_0$)	0.00	1.01	0.40	0.40
CV ($\%_0$)	15.81	64.20	40.68	40.68
FNR (%)	0.00	7.56	0.18	0.18
Cost	200	12.50	31.75	31.75
Intermediate parasitemia				
$\theta = 1,000$ parasites/ μ l				
Method	A	B	C	D
ME ($\%_0$)	0.00	1.01	0.40	0.40
CV ($\%_0$)	5.00	23.53	14.85	14.85
FNR (%)	0.00	0.00	0.00	0.00
Cost	200	12.50	31.75	31.75
High parasitemia				
$\theta = 10,000$ parasites/ μ l				
Method	A	B	C	D
ME ($\%_0$)	0.00	1.01	1.01	0.45
CV ($\%_0$)	1.58	14.03	14.03	9.31
FNR (%)	0.00	0.00	0.00	0.00
Cost	200	12.50	13	25.50

Measures of variability (ME, CV, FNR) and cost-effectiveness of methods are compared for fixed threshold values : 200 HPFs for Method A, 200 WBCs for Method B, 100 parasites and 200 WBCs for Method C, and 500 WBCs or 500 parasites for Method D.

doi:10.1371/journal.pone.0051987.t001

of *Cohen* [31] and its many variants and the *Scott's pi* [32] are the preferred statistics.

However, very few studies have examined the threshold-based counting techniques or evaluated the impact of the sampling error in endpoint measurements. In Nigeria, Dowling & Shute (1966) [12] showed that only 43% of infections in adults were detected by examining 200 fields, 61% by examining 600 fields and 70% by examining 1000 fields. In the Garki Project, Molineaux & Gramiccia (1980) [33] showed that the prevalence observed by the examination of 400 compared to 200 fields was increased by 10% for *P. falciparum*, by 24% for *P. malariae* and by 21% for *P. ovale*. Trape (1985) [11] compared the results of the examination of 100 and 200 fields of the thick film in 245 schoolchildren aged 6 to 16 from Linzolo (Congo). He concluded that the systemic examination of 200 oil immersion fields of the thick smear is the best compromise between the need for precision and rapidity. Prudhomme O'Meara et al. (2006) [34] showed empirically that counting beyond 200 WBCs may not significantly improve parasite density measurements.

In addition, the accuracy and consistency of these methods have been generally overlooked. There is no general agreement on the optimal method for estimating parasite density according to threshold values. Further experimental evidence is needed to determine which parasite counting technique is most accurate, reproducible, and efficient. Ultimately, the question is: to which extent would threshold values (specifically the number of WBCs counted and HPFs seen) influence the variability in parasite density estimates? However, there remains the issue of homoge-

neity. The distribution of the thickness of the smear and hence the distribution of parasites within the smear is not completely homogeneous [21]. Therefore, a proportion of the variability may be explained by this homogeneity factor.

To understand how the thresholds involved in parasite enumeration methods contribute to the magnitude of discrepancies in density determination, we studied their impact in variability measures generated by commonly used threshold-based counting techniques. We showed that estimators perform quite differently according to threshold values, and that an overall performance measure probably hides a lot of complexity in the behavior of each estimator. Another important aspect of this study is that we observed how estimators perform at different parasitemia levels, and how much the choice of threshold values may influence the performance of estimators relative to each parasitemia level. In summary, while all four estimators had some deficiencies, Method D outperformed all the other estimators for accuracy, precision measures and cost-effectiveness, and should therefore be seriously considered in future studies of comparative performance of PD estimators with field-collected data. In this paper, we explored the duality between cost-effectiveness and precision implied by estimation methods. An open question remains: To what extent is it possible to reduce methods' cost while staying accurate and precise in estimation measures?

Accurate estimation of PD is an important endpoint in epidemiological studies and clinical trials, both as a direct measure of the level of infection in a population and when defining parasitemia thresholds to diagnose malaria in case of fever

episodes. Malaria PD estimates are also used to assess the development of naturally acquired immunity [35] and in malaria vaccine investigations [23,36,37]. Therefore, inaccurate estimation of PD can lead to patient mismanagement and public health misinformation [38].

However, two approaches must be distinguished. The first approach concerns clinical malaria diagnosis and two problems can be pointed out depending on whether the question concerns an individual or a population. At the individual level, recent studies have highlighted the massive problem of misdiagnosis in malaria endemic countries [39–41]. From a clinical point of view the question is to determine whether or not a person presenting fever suffers from malaria or not and, in that sense, the main problem is a false negative result. Here, if a measure is falsely negative, the patient will be miscategorized and incorrectly treated, and measurement errors can lead to poor patient outcome. Nevertheless, although the level of infection is considered as a controversial sign of potential severity [42], treatment and medical supervision must be immediately started even if the PD is not accurately determined.

From an epidemiological point of view (e.g. to determine the incidence or prevalence of clinical malaria in an area or in a population under close medical surveillance) clinical malaria is often considered as any case of fever or fever-related symptoms (headache, vomiting, subjective sensation of fever) associated with a *P. falciparum* parasite/leukocyte ratio higher than an age-dependent pyrogenic threshold of PD previously identified in the patient [43,44]. In this case, a feverish individual harboring a PD under his age-specific threshold is not considered as a malaria case and will be monitored by the medical team involved in the study. In such a situation the accuracy of PD determination is obviously of great importance, not only for the patient but also for the outcomes and the conclusions of the study.

A second approach concerns the assessment of parasite density in epidemiological studies, when PD is used as the variable of

interest, independently of clinical disease. For example, genetic epidemiology studies often focused on a mean level of *P. falciparum* infection during a follow-up period [43,45–47]. Great care must also be taken in the analysis of parasite density estimates when parasite density is related to other explanatory variables, malariometric (e.g. parasite ratio, gametocyte ratio, mixed infection) or not (age, environmental or behavioral factors, medicine intake in clinical trials), when using statistical models as logistic regression and linear mixed effect models. In these cases as well as in population studies using a pyrogenic threshold to define clinical malaria, inaccurate estimates of parasite density might influence the parameters of associations between drug efficacy and the incidence of clinical malaria episodes in field trials [6], or between risk factors in epidemiological studies.

In further support of the arguments cited in this paper, empirical validation of the theoretical results is needed through a rereading experience conducted in the field. And toward a better understanding of threshold effects, we are interested in the study of the consequences of the quality of these estimators in models classically used and starting from these measures (mixed effects linear and logistic regression, generalized linear models, etc). These aspects of the problem are now under consideration.

Supporting Information

File S1 Supporting Information File.
(PDF)

Author Contributions

Technical Support: GN. Reviewed the manuscript: GN AG. Conceptual advice: AG. Discussed the results and implications and commented on the manuscript at all stages: IH GN AG. Conceived and designed the experiments: IH. Analyzed the data: IH. Wrote the paper: IH.

References

- Bejon P, Andrews L, Hunt-Cooke A, Sanderson F, Gilbert S, et al. (2006) Thick blood film examination for plasmodium falciparum malaria has reduced sensitivity and underestimates parasite density. *Malaria Journal* 5: 104.
- Christophers R (1951) Microscopic diagnosis of malaria. *British medical journal* 1: 75–76.
- Colbourne MJ (1971) The laboratory diagnosis of malaria. *Tropical Doctor* 1: 161–163.
- Collier JA, Longmore JM (1983) The reliability of the microscopic diagnosis of malaria in the field and in the laboratory. *Annals of Tropical Medicine and Parasitology* 77: 113–117.
- Draper CC (1971) Malaria. laboratory diagnosis. *British medical journal* 2: 93–95.
- Kilian A, Metzger W, Mutschelknauss E, Kabagambe G, Langi P, et al. (2000) Reliability of malaria microscopy in epidemiological studies: results of quality control. *Trop Med Int Health* 5: 3–8.
- WHO (1961) Manual for processing and examination of blood slides in malaria eradication programmes. WHORA1262-61.
- Armstrong Schellenberg J (1994) What is clinical malaria? finding case definitions for field research in highly endemic areas. *Parasitology Today* 10: 439–441.
- Warhurst D, Williams J (1996) Acp broadsheet no 148. July 1996. laboratory diagnosis of malaria. *J Clin Pathol* 49: 533–538.
- Moody A (2002) Rapid diagnostic tests for malaria parasites. *Clin Microbiol Rev* 15: 66–78.
- Trape JF (1985) Rapid evaluation of malaria parasite density and standardization of thick smear examination for epidemiological investigations. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 79: 181–184.
- Dowling MAC, Shute GT (1966) A comparative study of thick and thin blood films in the diagnosis of scanty malaria parasitaemia. *Bulletin of the World Health Organization* 34: 249–267.
- Garcia A, Dieng A, Rouget F, Migot-Nabias F, Le Hesran J, et al. (2004) Role of environment and behaviour in familial resemblances of plasmodium falciparum infection in a population of senegalese children. *Microbes and Infections* 6: 68–75.
- Kremsner P, Zotter G, Feldmeier H, Graninger W, Rocha R, et al. (1988) A comparative trial of three regimens for treating uncomplicated falciparum malaria in Acre, Brazil. *J Infect Dis* 158: 1368–1371.
- Le Port A, Cottrell G, Dechavanne C, Bouraima A, Briand V, et al. (2011) Prevention of malaria during pregnancy: assessing the effect of the distribution of iptp through the national policy in benin. *Am J Trop Med* 82: 270–275.
- Planche T, Krishna S, Kombila M, Engel K, Faucher JF, et al. (2001) Comparison of methods for the rapid laboratory assessment of children with malaria. *Am J Trop Med Hyg* 65: 599–602.
- WHO (2010) Basic Malaria Microscopy: Part I. Learner's Guide, Second Edition. World Health Organization.
- Earle WC, Perez M (1932) Enumeration of parasites in the blood of malaria patients. *Journal of Laboratory and Clinical Medicine* 17: 1124–1130.
- Chippaux JP, Akogbeto M, Massougboji A, Adjagba J (1991) Mesure de la parasitémie palustre et évaluation du seuil pathogène en région de forte transmission permanente. In: *Le paludisme en Afrique de l'Ouest : études entomologiques et épidémiologiques en zone rurale et en milieu urbain*, ORSTOM, Etudes et Thèses. pp. 55–65.
- Greenwood B, Armstrong J (1991) Comparison of two simple methods for determining malaria parasite density. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 85: 186–188.
- Alexander N, Schellenberg D, Ngasala B, Petzold M, Drakeley C, et al. (2010) Assessing agreement between malaria slide density readings. *Malaria Journal* 9: 4.
- Kirkwood B, Sterne J (2001) *Essentials of Medical Statistics*. Wiley-Blackwell, 2 edition.
- Petersen E, Marbiah NT, New L, Gottschau A (1996) Comparison of two methods for enumerating malaria parasites in thick blood films. *Am J Trop Med Hyg* 55: 485–489.
- Bland M, Altman D (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1: 307–310.
- Clendennen T, Long G, Baird K (1995) Qbc R and giemsa-stained thick blood films: diagnostic performance of laboratory technologists. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 89: 183–184.

26. Dubey M, Weingken C, Ganguly N, Mahajan R (1999) Comparative evaluation of methods of malaria parasite density determination in blood samples from patients and experimental animals. *Indian J Med Res* 109: 20–27.
27. Mulder B, van der Ligt W, Sauerwein R, Verhave P (1998) Detection of plasmodium falciparum gametocytes with the qbc R test and giemsa-stained thick blood films for malaria transmission studies in cameroon. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 92: 395–396.
28. Payne D (1988) Use and limitations of light microscopy for diagnosing malaria at the primary health care level. *Bulletin of the World Health Organization* 66: 621–626.
29. Prudhomme O'Meara W, McKenzie F, Magill A, Forney J, Permpanich B, et al. (2005) Sources of variability in determining malaria parasite density by microscopy. *Am J Trop Med Hyg* 73: 593–598.
30. Braun-Munzinger R, Southgate B (1992) Repeatability and reproducibility of egg counts of schistosoma haematobium in urine. *Trop Med Parasitol* 43: 149–154.
31. Fleiss JL, Cohen J (1973) The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement* 33: 613–619.
32. Scott W (1955) Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly* 17: 321–325.
33. Molineaux L, Gramiccia C (1980) *The Garki Project*. Geneva: World Health Organization.
34. Prudhomme O'Meara W, Remich S, Ogutu B, Lucas M, Mtalib R, et al. (2006) Systematic comparison of two methods to measure parasite density from malaria blood smears. *Parasitology Research* 99: 500–504.
35. Rogier C, Trape JF (1993) Malaria attacks in children exposed to high transmission: who is protected? *Transactions of the Royal Society of Tropical Medicine and Hygiene* 87: 245–246.
36. Alonso P, Smith T, Armstrong Schellenberg J (1994) Randomised trial of efficacy of spf66 vaccine against plasmodium falciparum malaria in children in southern tanzania. *Lancet* 344: 1175–1181.
37. Small DS, Cheng J, Hava TRT (2010) valuating the efficacy of a malaria vaccine. *The International Journal of Biostatistics* 6.
38. Bates I, Bekoe V, Asamo-Adu A (2004) Improving the accuracy of malaria-related laboratory tests in ghana. *Malaria Journal* 3: 38.
39. Amexo M, Tolhurst R, Barnish G, Bates I (2004) Malaria misdiagnosis: effects on the poor and vulnerable. *Lancet* 364: 1896–1898.
40. Reyburn H, Mbatia R, Drakeley C, Carneiro I, Mwakasungula E, et al. (2004) Overdiagnosis of malaria in patients with severe febrile illness in Tanzania: a prospective study. *BMJ* 329: 1212–1217.
41. Zurovac D, Midia B, Ochola S, English M, Snow R (2006) Microscopy and outpatient malaria case management among older children and adults in Kenya. *Trop Med Int Health* 11: 1–9.
42. WHO (2000) Spatial targeting of interventions against malaria. *Bull World Hlth Organ*.
43. Millet J, Nuel G, Watier L, Courtin D, Slaoui Y, et al. (2010) Genome wide linkage study, using a 250 k snp map, of plasmodium falciparum infection and mild malaria attack in a senegalese population. *PLoS One* 5.
44. Rogier C, Commenges D, Trape J (1996) Evidence for an age-dependent pyrogenic threshold of plasmodium falciparum parasitemia in highly endemic populations. *Am J Trop Med Hyg* 54: 613–619.
45. Garcia A, Cot M, Chippaux J, Ranques S, Feingold J, et al. (1998) Genetic control of blood infection levels in human malaria: evidence for a complex genetic model. *Am J Trop Med Hyg* 58: 480–488.
46. Sakuntabhai A, Ndiaye R, Casademont I, Peerapittayamongkol C, Rogier C (2008) Genetic determination and linkage mapping of plasmodium falciparum malaria related traits in senegal. *Microbes and Infections* 3: 2000.
47. Timmann C, Evans JA, König IR, Kleensang A, Rüschemdorf F, et al. (2007) Genome-wide linkage analysis of malaria infection intensity and mild disease. *PLoS Genet* 3: e48.