INTRODUCTION

Visual inspection of blood smears by light microscopy has been the standard diagnostic test for malaria for more than a century. It remains the least expensive and most widely used method. It is the only method available for quantifying parasitemia in infected blood. However, the sources and scale of error inherent in this technique have not been systematically investigated. Here we use data collected in outpatient clinics in Peru and Thailand to elucidate important sources of variation in parasite density measurements. We show that discrepancies between readings from two independent microscopists and multiple readings from a single microscopist are inversely related to the density of the infection. We present an example of how differences in reader technique, specifically the number of white blood cells counted, can contribute to the differences between readings. We discuss the implications of this analysis for field studies and clinical trials.

METHODS

Study participants. Study participants were enrolled during the high transmission seasons from among patients who came to local clinics in Iquitos, Peru and Maesod, Tak Province, Thailand. The protocols were reviewed and approved by the Ethical Committee for Research in Humans of the Thai Ministry of Public Health, the United States Army Medical Research and Materiel Command Human Subjects Research Review Board, the Committee for the Protection of Human Subjects at the Naval Medical Research Center, and the Comité de Ética, Instituto Nacional de Salud in Lima, Peru. The purposes and procedures of the study were explained in the ethnic dialect of the subjects, information was provided both orally and in written form, and informed consent was obtained from each subject prior to enrollment. An independent witness also signed each consent form. Subjects less than 18 years of age in Peru were enrolled after consent was obtained from the individual and their legal guardian. Patients with fever, or a history of fever within the past 72 hours, and no treatment with anti-malarial drugs within the past two weeks were enrolled in the study. Adults and children more than one year of age were enrolled in Peru between June 23 and August 17, 1998 and between May 17 and June 19, 1999. Adults more than 15 years of age were enrolled in Thailand between May 28 and August 28, 1998 and between June 7 and July 9, 1999. Details of the study have been described previously.

Sample collection. Methods for collecting samples, preparing slides, and reporting parasite species and densities have been previously reported in detail. Briefly, 2–4 mL of blood were collected by venipuncture from each participant. From this sample, three slides, each with a thick and thin film, were prepared. For the thick smears, 6 μL of blood were micropotted onto clean slides; for the thin smears, 4 μL of blood were used. The clinic staff used one slide immediately for diagnosis and treatment purposes. The remaining two slides (slide no. 1 and slide no. 2) were stained with Giemsa following standard procedures. Some of the remaining blood sample from each patient was used to determine the WBC count per microliter of blood, using automated cell counters.

Microscopy. Microscopists used identical model microscopes with identical type objective lenses. Slide no. 1 was read by two skilled microscopists (A and B), each blinded to the results reported by the other. Slide no. 2 was reserved for review to aid in determining a final study result when the readings from microscopists A and B were discordant. The microscopists counted the number of parasites per WBCs in consecutive oil-immersion high-power fields until the field

Abstract. Enumeration of parasites by microscopic examination of blood smears is the only method available for quantifying parasitemia in infected blood. However, the sources and scale of error inherent in this technique have not been systematically investigated. Here we use data collected in outpatient clinics in Peru and Thailand to elucidate important sources of variation in parasite density measurements. We show that discrepancies between readings from two independent microscopists and multiple readings from a single microscopist are inversely related to the density of the infection. We present an example of how differences in reader technique, specifically the number of white blood cells counted, can contribute to the differences between readings. We discuss the implications of this analysis for field studies and clinical trials.
1. REPORT DATE
2005

2. REPORT TYPE
N/A

3. DATES COVERED
-

4. TITLE AND SUBTITLE
SOURCES OF VARIABILITY IN DETERMINING MALARIA PARASITE DENSITY BY MICROSCOPY

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Fogarty International Center, National Institutes of Health, Bethesda, Maryland; Walter Reed Army Institute of Research Silver Spring, Maryland

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Fogarty International Center, National Institutes of Health, Bethesda, Maryland; Walter Reed Army Institute of Research Silver Spring, Maryland

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Medical Research Center 503 Robert Grant Avenue Silver Spring, MD 20910-7500

10. SPONSOR/MONITOR’S ACRONYM(S)

11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:
   a. REPORT unclassified
   b. ABSTRACT unclassified
   c. THIS PAGE unclassified

17. LIMITATION OF ABSTRACT SAR

18. NUMBER OF PAGES 6

19a. NAME OF RESPONSIBLE PERSON

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18
with the 200th WBC was reached and counted. If malaria parasites were observed, but the total number of parasites counted at the 200 WBC mark was less than 10, counting of parasites and WBCs in consecutive oil-immersion high-power fields continued until the field with the 500th WBC was reached and counted. Since all parasites and WBCs were counted in the final field, the exact number of WBCs counted and recorded was often slightly greater than 200 or 500. The exact number of WBCs counted, rather than a rounded approximation of 200 or 500 WBCs, was used in the calculation of parasite density. Two hundred oil-immersion high-power fields were examined on the thick film before a slide was declared negative for malaria parasites. In cases in which microscopists A and B disagreed about the presence of parasites, disagreed about the species present, or reported densities that differed by more than a factor of 2, a senior microscopist read both slide no. 1 and slide no. 2. For parasitemias greater than \(100,000/\mu\text{L}\), readings that differed by more than \(50,000/\mu\text{L}\) were also considered discordant and read by the senior microscopist. In addition, a sample of 5% of the slides in which microscopists A and B had concordant readings (whether positive or negative for parasites) was read by microscopist C as a quality control measure. For each microscopic reading, the total number of asexual parasites counted on the slide was divided by the total number of WBCs counted, then multiplied by the measured WBC count for the patient (rather than by a standard approximate WBC count for all patients). The result was reported by each microscopist as parasites per microliter of whole blood.

**Analysis.** We eliminated from the analysis all samples that had been identified as mixed-species infections (8 and 13 from Peru in 1998 and 1999, respectively, and 15 and 17 from Thailand in 1998 and 1999, respectively), or for which any two (A, B, or C) microscopists disagreed on the species present (32 and 37 from Peru in 1998 and 1999, respectively, and 57 and 18 from Thailand in 1998 and 1999, respectively). When comparing the results of microscopists A and B, we disregarded samples in which parasites were observed by only one microscopist. A total of 1,911 samples were read as parasite positive by two microscopists and included in this analysis (550 from Peru and 1,361 from Thailand). In 242 instances (98 from Peru and 144 from Thailand), microscopists A and B differed in their density estimates by more than two-fold; these were read by a third microscopist.

**RESULTS**

**Variability between readers.** Two microscopists (A and B) examined a single slide with a thick and a thin blood film, and, if parasites were detected, reported the species and the number per microliter. Considering only the single-species infections, without regard for species identification, the scaled difference between the parasite densities reported by A and B (difference between A and B divided by the mean of A and B) is plotted versus the mean parasite density in Figure 1. Discrepancies in reported densities decrease with increasing mean density. The sign of the Spearman rank test statistic for correlation between discrepancy and density was negative and correlation was significant for each site and each year \((P < 0.0001\) for all except Peru 1999, where \(P = 0.0016\)). Fifty-percent of samples in this analysis from Thailand in 1998 and 1999 were *Plasmodium falciparum* compared with only 36% and 18% from Peru in 1998 and 1999, respectively. Trends between density and discrepancy were similar when *P. falciparum* and *P. vivax* were considered separately and the correlation between discrepancy and density was significant in all cases (Spearman \(P < 0.001\), except for *P. falciparum* in Peru in 1998 \((P = 0.065)\) and *P. vivax* in Peru in 1999 \((P = 0.073)\). The median scaled difference for the Thai data (both years and both species) was 0.16, and the median density was 4,672 parasites/\(\mu\text{L}\). The median difference and density for the Peruvian data were 0.19 and 5,520, respectively.

**Variability due to technique.** Microscopists counted parasites per WBC until the field with the 200th WBC was counted; if at that point the number of parasites counted was less than 10, they continued counting until the field with the 500th WBC. Therefore, each slide has two parasite counts associated with it, one from each reader, and two WBC counts, one from each reader. If both readers counted until 200 WBCs and stopped, then the total WBCs counted by the microscopists for that slide was approximately 400. If both counted until 500 WBCs, the total was approximately 1,000, but if one stopped at 200 WBCs and the other reader continued until 500 WBCs, the total WBCs counted was approximately 700. In this last case, the technique of the two readers diverged in a single, measurable way. To determine if discrepancies between the densities reported from the same slide by two different readers are associated with the number of WBCs counted by the microscopists, for each of these instances we plotted the scaled difference between the measurements from each microscopist versus the total number of

![Figure 1](image-url)
WBCs counted for that slide. The magnitude of discrepancies was not proportional to the number of WBCs indexed by the readers. Instead, the greatest discrepancies were seen when the two reported parasite densities were based on different WBC counts. Figure 2 shows the data from Thailand in 1999; similar trends are evident in the data from both countries in both years. Table 1 shows the median scaled differences between parasite densities for samples counted up to 400, 700, or 1,000 total WBCs and the 95% confidence intervals around the median. The median differences are significantly higher (Mann-Whitney P < 0.001) for the parasite densities with discordant numbers of WBCs counted by each reader. Differences for readings based on consistent counts of 500 WBC per reader (1,000 WBCs total) are higher than those for 200 WBC per reader (400 WBCs total) because, by the nature of the procedure, the median parasite density for samples counted to 1,000 WBCs is much lower (61/μL versus 7,287/μL). Similarly, samples for which 700 WBCs were counted had a lower median density than samples counted to 400 WBCs because one of the two microscopists had observed less than 10 parasites after 200 WBCs had been counted (356 parasites/μL versus 7,287 parasites/μL). We showed earlier that discrepancies increase with decreasing density. To eliminate density as a variable and to compare directly between samples of similar densities that were counted to different numbers of total reference WBCs, we focused on samples with mean densities between 200 and 400 parasites per microliter. Samples with densities in this interval were present in all three WBC count categories (400, 700 or 1,000 total WBCs counted). The median scaled differences in readings were nearly identical for the 400 and 1,000 WBC count categories (median = 0.16 and 0.2, respectively, Mann-Whitney P = 1.0), but was more than four times higher for the 700 WBC count category (median discrepancy = 0.86, Mann-Whitney P < 0.0001). This analysis was complicated by the fact that a discrepancy already existed at the 200 WBC mark (one reader had observed < 10 and the other ≥ 10 parasites per 200 WBCs; the former continued counting and the latter stopped), but it is not clear how (or if) this discrepancy contributes to the results discussed earlier.

Variability between slides. A total of 242 matched pairs of slides, each pair from a single patient, were read by a quality control microscopist (C). It is important to note that the person performing this function within a study site was not necessarily the same individual for all the samples in 1998, but a single individual at each site performed the microscopist C function in 1999. In Figure 3, the scaled difference between the parasite densities reported from the two slides (slide no. 1 – slide no. 2 divided by the mean parasite density of slide no. 1 and slide no. 2) is plotted as a function of the mean density of the two slides. Variation between readings from two slides from the same patient is also density dependent. Discrepancies between the two readings of microscopist C were lower in Thailand (median difference = 0.08) than in Peru (median difference = 0.17, Mann-Whitney P < 0.001). The median per-microliter parasite density on slides read by microscopist C was 439 in Thailand versus 985 in Peru. Although both the median density and median discrepancy were lower in Thailand, there was a significant inverse correlation between density and discrepancy between slides for each site and each year. In Thailand, discrepancies between slides were significantly lower than discrepancies between readers (Mann-Whitney P < 0.001), but in Peru the differences between readers and between slides could not be distinguished (Mann-Whitney P = 0.936). Figure 4 shows the frequency histogram of the percentage difference between slides read by microscopist C color-coded according to the mean density (evaluated from the two counts recorded by microscopist C) of the infection, with infections ≤ 500 parasites/μL categorized as low density, those of 500–5,000 parasites/μL as medium density and those > 5,000 parasites/μL as high density. For 30% (Peru) to 40% (Thailand) of paired slides, reported parasite densities differed from each other by 1–10%, with lower-density slides showing greater differences; there were up to two-fold differences between estimates by the same microscopist from the two slides.

**DISCUSSION**

Accurate determination of parasite density is an essential clinical trial endpoint, both as a direct measurement and when using a threshold parasitemia to define an episode of malaria. It is also of value in clinical settings in high-transmission areas when determining the cause of a febrile episode. Errors inherent in diagnosis with microscopy, including those associated with sample preparation, staining, counting, and technician performance, have been reported for at least 75 years, but the accuracy and consistency of the technique have been generally taken for granted. The lack of evidence for the validity of these assumptions is striking.

To determine the magnitude of discrepancies in density determination in a setting consistent with clinical trial procedures (i.e., samples collected and stained in the field and read by trained microscopists), we used a large dataset collected in
two years at sites on two continents, and analyzed the parasite densities reported from 1,911 positive, single-species infections. Parasite counts were done uniformly, with the number of WBCs indexed by the microscopist dictated by the protocol, thus eliminating a potential source of variation. In addition, WBC counts per microliter were measured from whole blood for each patient, allowing true parasite densities per microliter to be determined and thus permitting trends based on density to be analyzed accurately. If each count had been multiplied by a uniform approximation of WBCs per volume, as is usually done with microscopy data, the true relationship between density and discrepancy would have been obscured.\textsuperscript{10}

The median discrepancy observed when two microscopists examined the same slide was 15–17%, with a strong dependence on parasite density. The differences between measurements could be attributed to differences in individual reader technique and discretion in identifying parasites or to random sampling error due to the distribution of parasites on the slide and differences in fields read by the two microscopists.\textsuperscript{11} If the error was due entirely or in part to the latter, then for samples of similar density, the discrepancy would be expected to decrease as the number of WBCs counted (i.e., number of fields examined) increased. However, in samples with densities of 200–400 parasites/μL, differences were not inversely proportional to the number of WBCs counted. Densities reported based on discordant numbers of WBCs had significantly higher discrepancies between the two readings. One conclusion from this observation is that identical, consistent technique may be more important than increasing the number of fields read when precision (agreement between readings) among readers is desired, a result that may expedite the collection of data and reduce reader workload and fatigue. In this study, instructing readers to stop after the field containing the 200th WBC, regardless of parasite density, might have resulted in greater agreement between readers than the 200/500 WBC scheme used here. Although this modification may improve precision, it is not clear which estimate would be closer to the true density in a peripheral blood sample, and it is therefore not certain that accuracy would improve. This suggestion requires further investigation due to the unknown effect of the difference between readings that necessarily existed when one reader stopped after the field containing the 200th WBC and the second reader continued through the field with the 500th WBC. Although we cannot dissect the error introduced by this variable stopping scheme of counting, it seems unlikely that this factor alone could account for the increased discrepancy between measurements from accordant and discordant WBC counts. It is noteworthy that in a study by Kilian and others,\textsuperscript{4} differences between two readers who counted parasites per volume rather than parasites per WBC showed the same decreasing trend with increasing density. However, the mean difference between the readers was 9%, somewhat lower than the median difference of 15–17% reported here. Discrepancy will certainly depend on the specific protocol (i.e., the number of WBCs or fields counted, etc.), therefore, we agree with Hanscheid and Valadas,\textsuperscript{12} who urged that all studies include an internationally recognized standard method for reading slides. We add that detailed protocols should always be reported.

Finally, to estimate the differences between slides prepared from the same patient, we plotted discrepancy as a function of density for 242 instances in which two slides were read by a single microscopist. For these 242 samples, original readings by microscopists A and B differed by more than two-fold. In the data from Peru, the discrepancy between slides was equivalent in magnitude to the discrepancy between readers. This could indicate that the readers were as good as the slides;
inconsistencies in slide thickness and parasite distribution in the film were greater than differences in reader technique. In the Thai data, differences between reads from the same slide (different microscopists) were greater than differences between reads from different slides (same microscopist). This could be interpreted to mean that the differences between readers were due to reader technique and could be improved by further training.

These results highlight important considerations for trials with density endpoints. Errors in density measurements are not constant, but are a function of density and include error due to reader technique, slide quality and the random distribution of parasites and WBCs within the blood film. Sources of error to be addressed and evaluated in a systematic manner include estimates of parasite loss during sample handling and staining, degradation during storage, experimental error arising from counting method, and random error associated with the statistical distribution of parasites and WBCs in the blood smear. Some of these sources of error are not likely to be reduced by training or standardization of technique, but cannot be isolated from the current dataset. Other sources of variation in density data that fall outside the scope of microscopy technique, but could be particularly important for vaccine trials and merit careful consideration, are the temporal variation in peripheral parasitemia caused by parasite synchronization and sequestration.\(^\text{13,14}\) Sequential samples taken from each patient during a study could potentially improve the sensitivity and accuracy of efficacy estimates.

Received February 25, 2005. Accepted for publication April 18, 2005.

Acknowledgments: The authors thank Dr. J. Sirichaisinthop for his contributions to this work. We also thank the Armed Forces Research Institute of Medical Sciences and the Naval Medical Research Center Detachment field study teams for their dedication and technical assistance.

Financial support: This work was supported by the United States Army Medical Research and Materiel Command.

Disclaimer: The views presented are those of the authors, and do not necessarily reflect the official policy or position of the U.S. Departments of the Navy or Army, the U.S. Department of Defense, the U.S. Government, or any other organization listed.

Authors' addresses: Wendy Prudhomme O’Meara and F. Ellis McKenzie, Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, E-mails: prudhomw@mail.nih.gov and mckenzell@mail.nih.gov. Alan J. Magill, J. Russ Forney, and Robert A. Gasser, Jr., Walter Reed Army Institute of Research, Silver Spring, MD 20910, E-mails: Alan.magill@na.amedd.army.mil, John.Forney@usma.edu, and rgasser@erols.com. Barnyen Permpanich, Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand. Carmen Lucas, United States Naval Medical Center Detachment, Lima, Peru, E-mail: clucas@nmrcd.med.navy.mil. Chan- suda Wongsrichanalai, United States Naval Medical Research Unit No. 2, Jakarta, Indonesia, E-mail: chansuda@namru2.org.

REFERENCES


