DEPENDENCE OF MALARIA DETECTION AND SPECIES DIAGNOSIS BY MICROSCOPY ON PARASITE DENSITY

F. ELLIS MCKENZIE, JEERAPHAT SIRICHAI SINTHOP, R. SCOTT MILLER, ROBERT A. GASSER JR., AND CHANSUDA WONGSIRICHANALAI

Fogarty International Center, National Institutes of Health, Bethesda, Maryland; Department of Communicable Disease Control, Ministry of Public Health, Bangkok, Thailand; Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand; Walter Reed Army Institute of Research, Silver Spring, Maryland

Abstract. Giemsa-stained blood smears from each of 2,190 patients from Thai government-operated clinics on the Thailand-Myanmar border were independently examined by the on-duty microscopists at the clinics and by 2–3 research microscopists, each blinded to the clinics’ and each other’s reports. Using a strictly defined protocol, a consensus reference-standard blood smear interpretation for each sample was produced by the research microscopists. This result was compared with the clinic’s diagnostic interpretation for the corresponding sample with respect to detection of parasitemia and diagnosis of infecting species. Reference-standard results reported parasitemia in 13.2% of the samples reported negative by the clinic. Reference-standard results were negative in 24.3% of the samples reported parasite-positive by the clinic. For samples in which both the reference-standard result and the clinic result reported parasitemia, species identification differed for 13.7% of the samples. The likelihood of parasite detection and correct diagnosis at the clinic varied in accordance with the reference-standard estimates of parasite density.

INTRODUCTION

Effective malaria treatment depends on prompt, accurate detection and diagnosis. Failures can lead to omission of a drug when a drug is required, administration of a drug when no drug is required, or administration of an ineffective drug. Non-rational drug use, in turn, can promote drug resistance. Studies of epidemiology and immunity depend on accurate detection, diagnosis, and density estimation. Failures confound attempts to evaluate the effects of interventions, on an individual-patient or population-wide basis. Thus, it is striking that, in relative terms, so much effort has been devoted to possible interventions, and so little to the methodology on which their evaluation depends, methodology that has barely changed over the past 100 years.

It is well known that the details of microscopic methods are important for detection and density estimation.1–4 Comparative testing of microscopists has a long history, some of which is reflected in the scientific literature.5,6 In particular, comparisons of clinic and reference microscopy indicate that mixed-species infections and Plasmodium falciparum gametocytes are often overlooked. In India7 and Uganda,8 clinic microscopy missed 93% and 41% of the P. falciparum gametocyte-positive samples detected by reference microscopy, respectively; in 100% and 46% of the mixed-species infections detected by reference microscopy, respectively, clinic microscopy reported only one species. Furthermore, discrepancies between the results of microscopic and molecular techniques suggest that many low-density asexual-form single-species infections also go undetected or misdiagnosed.9,10 Here we report comparative data from clinic and reference microscopy from a large study on the Thailand-Myanmar border to investigate the relationship of parasite densities to malaria detection and species identification in general, as well as the detection of gametocytes and diagnosis of mixed-species infections.

MATERIALS AND METHODS

Microscopy data were collected during a study of malaria rapid diagnostic devices, for which the methods are described in detail elsewhere.11,12 Briefly, participants were individuals ≥15 years old presenting to local malaria clinics at Maesod (on the Thailand-Myanmar border) between May 28 and August 28, 1998 or between June 7 and July 9, 1999, with a fever (oral temperature ≥38°C), headache, or self-reported history of fever within the previous 72 hours. The age distribution of the participants was relatively narrow (60% were 18–25 years old) because the majority of volunteers were migrant workers from Myanmar seeking agricultural employment in Thailand; we found no age-related effects in the results reported here. The study protocol was reviewed by the Human Use Review Committee, Walter Reed Army Institute of Research, and the Human Subjects Research Review Board, Office of the Surgeon General, and approved as Walter Reed Army Institute of Research Protocol #687. The study protocol was also approved by the Thai Ministry of Public Health Ethical Review Committee for Research in Human Subjects and implemented under the guidance of the Vector-Borne Disease Control Regional Office No. 1 (Phra buddhat, Thailand).

The clinic and Armed Forces Research Institute of Medical Sciences (AFRIMS) slides were made from the same draw of venous blood. For each individual, precise volumes of well-mixed whole blood were micropipetted and prepared as a thick smear on one slide, and a thick and a thin smear on each of two other slides. Smears were prepared according to a standardized method (blood volume measured, diameter of the thick film controlled) by a team of well-trained AFRIMS technicians. The thick-film-only slide was handed immediately to the clinic staff for drying and staining with Giemsa according to routine clinic procedures; this slide was the basis of the clinic’s diagnosis and treatment.

In accord with standard AFRIMS procedure, as is highly recommended for anticoagulated blood (in this case with EDTA), the other slides were held overnight before staining. Each of two AFRIMS microscopists (“A” and “B”), blinded to the other’s interpretation, read slide 1 of the other two slides. Two hundred oil-immersion high-power fields on the thick film were read before any slide was interpreted as negative; the thin film was used for species determination. If the A and B results disagreed on the presence or species of parasites, or by a factor of two or more on parasite density, further
readings were made by a senior microscopist, using both slide 1 and slide 2 ("A" and "B" results). Additionally, in 5% of cases in which the A and B results were in agreement, the senior microscopist reviewed both slides as a quality control measure. Clinic and AFRIMS diagnoses were based solely on parasite asexual blood forms.

Microscopy results were available from the AFRIMS (diagnosis and per microliter density estimates) for 3,298 individuals. Of these, microscopic results from the clinic (diagnosis only: uninfected, P. falciparum, P. vivax, mixed P. falciparum-P. vivax) were available for 2,190 (66.4%) individuals. The original protocol did not include the comparisons reported here, and so did not require clinic microscopy data. A protocol amendment was prepared and approved approximately three years later; when we obtained the clinic records, those from the earliest part of the study were no longer available. To the best of our knowledge, the missing clinic records had no special characteristics that might bias the analyses here.

We initially separated the 29 individuals for whom the A and B diagnoses disagreed, and the 24 in which the A, B, or clinic diagnoses were of a mixed-species infection. For each of the remaining 2,137 individuals, we calculated the mean and the coefficient of variation (CV) of the asexual-form density estimates reported by A and B. For each year, for each species, we compared distributions of the mean A and B density estimates among three categories: those in which the clinic and AFRIMS diagnoses agreed (true positive), those in which they agreed that an infection was present but disagreed on the species (species shift, i.e., from the clinic to the AFRIMS diagnosis), and those in which AFRIMS diagnosed that an infection was present and the clinic that infection was absent (false negative). We also compared the mean C and D asexual-form density estimates in which A and B diagnoses agreed to those in which the A and B diagnoses disagreed, and the asexual-form and gametocyte density estimates for readings in which gametocytes were detected to those in which they were not. For these comparisons of distributions we used the Kolmogorov-Smirnov two-sample test, and for contingency tables, we used the G-test.13 Tables 1–5 refer to individuals for whom the two AFRIMS diagnoses agreed.

AFRIMS maintains a stringent internal quality assurance process of malaria smear reading, which includes extensive training and periodic cross-checking by an investigator (CW) and a senior microscopist. With respect to the consistency of the AFRIMS readings, in only 11 (0.5%) of 2,166 single-species instances (i.e., P. falciparum or P. vivax, not both) was one AFRIMS microscopist’s diagnosis “uninfected” and the other’s “infected;” in only 18 (0.8%) did the two microscopists disagree about the species present. If instances in which disagreement was based solely on the presence or absence of gametocytes are included, these figures increase to 15 and 23, respectively.

In two of the 29 instances in which the first 2 AFRIMS readings (A and B) disagreed with respect to diagnosis, the referee microscopist’s diagnosis of slide 1 (C) differed from that of slide 2 (D). In 62 instances the first two AFRIMS readings agreed with respect to diagnosis but disagreed by a factor of two or more on parasite density; in two of these, the second readings agreed with each other on diagnosis, but disagreed with the first readings. For readings in which the consensus C, D diagnosis was a single-species infection, comparisons of the asexual-form density estimate distributions between the C, D readings for which the A, B diagnoses agreed and those for which they disagreed yielded P values >0.2, which provides increased confidence in the consistency of the AFRIMS density estimates.

Secondary AFRIMS readings (C and D) were available for five of the 14 instances in which primary diagnoses (A and B) of mixed-species infection agreed as well as the four in which they disagreed; for the four cases in which the primary readings disagreed, the secondary readings reported single-species infections in two and mixed-species infections in two. For these nine cases, comparisons of distributions of asexual-form density estimates between the A, B and C, D readings and between those for which the A, B diagnoses agreed and those for which they disagreed produced P values >0.08, which again increases confidence in the consistency of the AFRIMS density estimates.

### RESULTS

Table 1 shows that for 1998, 10.6% (57 of 536) of the individuals diagnosed by the clinic as uninfected were diagnosed by AFRIMS as infected; the corresponding figure for 1999 was 15.7% (89 of 568). For 1998, 25.5% (127 of 498) of the individuals diagnosed by the clinic as infected were diagnosed by AFRIMS as uninfected; the corresponding figure for 1999 was 23.2% (124 of 535). Thus, overall, the frequency of clinic false negatives was (57 + 89)/(536 + 568) = 146 of 1,104 = 13.2%, and that of clinic false positives was (127 + 124)/(498 + 535) = 251 of 1,033 = 24.3%. If instances in which the two AFRIMS readings disagreed are included, then each clinic false-negative percentage is slightly higher, and each clinic false-positive percentage is slightly lower. The overall frequency of discrepancies in species identification is calculated by dividing the total on which the clinic and AFRIMS diagnoses disagreed (107) by the total on which both agreed on parasites’ presence (782); thus 107 of 782 = 13.7%.

As shown in Table 2, the mean density estimates for P. vivax were always lower than those for P. falciparum. For both species, in both years, the mean density estimates for true positive cases were always greater than those for species-shift cases, which were always greater than those for false-negative cases. The low CV values indicate good agreement.
Mean *Plasmodium* asexual form density estimates per microliter

<table>
<thead>
<tr>
<th></th>
<th><strong>Estimated density</strong></th>
<th><strong>95% confidence limits</strong></th>
<th><strong>Coefficient of variation</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P. vivax</td>
<td>P. falciparum</td>
<td>P. vivax</td>
</tr>
<tr>
<td>1998</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True positive</td>
<td>3,020</td>
<td>10,000</td>
<td>2,993–3,047</td>
</tr>
<tr>
<td>Species shift</td>
<td>1,445</td>
<td>2,042</td>
<td>1,396–1,494</td>
</tr>
<tr>
<td>False negative</td>
<td>79</td>
<td>204</td>
<td>23–136</td>
</tr>
<tr>
<td>1999</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True positive</td>
<td>3,802</td>
<td>14,454</td>
<td>3,772–3,832</td>
</tr>
<tr>
<td>Species shift</td>
<td>1,288</td>
<td>1,995</td>
<td>1,257–1,320</td>
</tr>
<tr>
<td>False negative</td>
<td>200</td>
<td>331</td>
<td>162–237</td>
</tr>
</tbody>
</table>

between the AFRIMS microscopists. With only one exception, the mean CV for *P. vivax* is greater than that for *P. falciparum*. For 1999, for both species, the mean CV for true positive cases is less than that for the species-shift cases, which is less than that for the false-negative cases; with only one exception, 1998 shows the same pattern.

For readings in which the consensus AFRIMS diagnosis was a single-species infection, we compared the distributions of asexual-form density estimates for each year, for each species, among three categories: those in which the clinic and AFRIMS diagnoses agreed (true positive), those in which they agreed that an infection was present but disagreed on the species (species shift), and those in which AFRIMS diagnosed that an infection was present and the clinic that infection was absent (false negative). As shown in Table 3, all comparisons except those between the true *P. vivax* positives and the corresponding species shifts (i.e., those for which the clinic diagnosis was *P. falciparum* and the AFRIMS diagnosis was *P. vivax*) yielded *P* values of 0.001 or less. Taking the AFRIMS data as definitive, this indicates that the likelihood of parasite detection increases with density, and that the likelihood of correct species diagnosis does the same.

As noted earlier in this report, several previous, smaller studies in other regions indicated that *P. falciparum* gametocytes and mixed-species infections were often missed, so we included these in our analyses to the extent possible. It is not usual clinic practice to take note of gametocytes, and in fact they were noted in clinic diagnoses only in 1999, only for *P. falciparum*, and only in 16 instances, 11 of which were confirmed by both AFRIMS microscopists; in three instances the AFRIMS microscopists disagreed with each other. Both AFRIMS microscopists reported *P. falciparum* gametocytes in 24 other instances in 1999; in another seven, only one of the two AFRIMS microscopists reported *P. falciparum* gametocytes. The instances in which gametocytes were noted at the clinic were related to gametocyte density (*P* = 0.0025 for true positives versus false negatives), not to asexual-form density (*P* > 0.5). Both AFRIMS microscopists reported *P. vivax* gametocytes in 138 instances in 1999; in 23, only one of the two AFRIMS microscopists reported *P. vivax* gametocytes.

For 1998, both AFRIMS microscopists reported *P. falciparum* gametocytes in 14 instances; in another 13, only one of the two AFRIMS microscopists reported *P. falciparum* gametocytes. Both AFRIMS microscopists reported *P. vivax* gametocytes in 99 instances in 1998; in 40, only one of the two AFRIMS microscopists reported *P. vivax* gametocytes. Thus, overall for slides on which both AFRIMS readings reported gametocytes, 6.8% of *P. falciparum* and 46.0% of *P. vivax* infections were gametocytemic in 1998; the corresponding numbers for 1999 were 14.5% and 54.5%, respectively. Including the slides in which only one of the microscopists reported gametocytes increases these numbers to 13.2%, 64.6%, 18.7%, and 63.6%, respectively.

We compared the asexual-form and gametocyte density estimates for AFRIMS readings in which gametocytes were detected to those in which they were not (Table 4). As shown in Table 5, the asexual-form densities differ only for *P. vivax*, and only for gametocyte-positive versus gametocyte-negative comparisons (i.e., *P. vivax* gametocyte-positive specimens had significantly higher asexual-form parasite density than *P. vivax* gametocyte-negative ones); there were no clear patterns in the gametocyte-density comparisons.

Gametocytes were only rarely reported in the absence of asexual forms of the same species, but these reports were disproportionately associated with diagnostic disagreements and putative mixed-species infections. For instance, had gametocytes been included as a basis for diagnosis, the AFRIMS diagnoses of single-species infections would have increased by only 0.2% (5 of 2,123), where the two diagnoses agreed, but by 16% (4 of 25) where they disagreed. The corresponding figures for mixed-species infections are such that

<table>
<thead>
<tr>
<th></th>
<th><strong>P. vivax</strong></th>
<th><strong>P. falciparum</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Asexual</td>
<td>Gametocyte</td>
</tr>
<tr>
<td>1998 gametocyte status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive, both A and B</td>
<td>5,623</td>
<td>229</td>
</tr>
<tr>
<td>Positive, either A or B</td>
<td>3,802</td>
<td>138</td>
</tr>
<tr>
<td>Negative, both A and B</td>
<td>257</td>
<td>–</td>
</tr>
<tr>
<td>1999 gametocyte status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive, both A and B</td>
<td>5,129</td>
<td>186</td>
</tr>
<tr>
<td>Positive, either A or B</td>
<td>2,570</td>
<td>43</td>
</tr>
<tr>
<td>Negative, both A and B</td>
<td>339</td>
<td>–</td>
</tr>
</tbody>
</table>
these diagnoses would have increased by 7% (1 of 14) and 125% (5 of 4), respectively.

If the data in Table 1 are restructured as separate contingency tables for the clinic and AFRIMS in each year, G-test results for each table indicate a marked deficit of mixed-species infections (all P values <10^{-6}). Disagreements about putative mixed-species infections, between the clinic and AFRIMS or between the two AFRIMS readings, were always related to whether one or two species were present in an infection, never whether the individual was infected or not. Overall, in the 14 cases in which the two AFRIMS microscopists agreed on a mixed-species diagnosis, the mean asexual-stage density estimates for \textit{P. falciparum} were higher and the \textit{P. vivax} estimates lower than in single-species cases.

<table>
<thead>
<tr>
<th>Year</th>
<th>Status</th>
<th>Positive A and B vs A or B</th>
<th>Positive A and B vs negative</th>
<th>Positive A or B vs A and B</th>
<th>Positive A or B vs negative</th>
<th>\textit{P. falciparum} Asexual</th>
<th>\textit{P. falciparum} Gametocyte</th>
<th>\textit{P. vivax} Asexual</th>
<th>\textit{P. vivax} Gametocyte</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>Gametocyte status</td>
<td>&gt;0.5</td>
<td>0.1</td>
<td>0.45</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>Gametocyte status</td>
<td>10^{-12}</td>
<td>–</td>
<td>0.325</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10^{-10}</td>
<td>–</td>
<td>&gt;0.5</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textbf{DISCUSSION}

Our results demonstrate that parasite densities determine both malaria detection and species identification in a malaria clinic on the Thailand-Myanmar border. These phenomena are surely much more widespread. We know of no previous work that addresses species identification, but with respect to malaria detection, two classic studies showed rising frequencies of positive samples as more fields\textsuperscript{14} or more films\textsuperscript{15} were examined per patient. The Thai malaria control program is known to have one of the most well-established clinic microscopy systems. Nonetheless, based on the AFRIMS reference standard, these data showed that clinic microscopy had false-negative rates exceeding 10%, and false-positive rates more than 20%.

Clinic microscopists are advised to read 100 (rather than 200) oil-immersion high power fields per thick film, but in practice this goal may not be achievable depending on individual microscopist motivation and workload. This study was conducted during peak malaria seasons, when their workloads were heaviest (>100 films/microscopist/day). Giemsa staining could have been poorly standardized. High humidity during the rainy days of the peak season could be responsible for poor fixing of blood films. While these factors may account for the high frequencies of false-positive and false-negative results, and contribute to misdiagnoses in general, they would not explain why misidentification of \textit{P. falciparum} as \textit{P. vivax} was density-dependent, while the reverse was not. The frequency of false-positive readings by malaria clinics in this study seemed unacceptable, though limited specificity (71%) of basic malaria microscopy had previously been noted.\textsuperscript{16} At this point, of course, it is impossible to know whether the false-positive results were platelets, red blood cell fragments, staining artifacts, or something else. These problems will be further explored, and efforts to improve diagnostic capability at the clinic are underway.

It is possible that the relatively poor clinic performance reported in this paper is specific to the individual malaria clinic during the two years of this study. Performance of clinic microscopists that seems adequate for a malaria control program that includes active case-finding and follow-up, residual spraying, and other control measures is still likely to vary greatly and should be more closely monitored. A quality-control process exists at Thai malaria clinics but may need to be improved, for instance by increasing the percentage of slides that is randomly selected for cross-checking at a Zone/Regional Office (currently only up to 10%) and organizing more frequent refresher courses for microscopists in remote clinics. Recent budgetary cuts and downsizing of the government workforce (such that retired expert microscopists are not replaced) may have indirectly lead to weaker quality control.

We assume that detection and diagnosis would improve if clinic microscopists spent more time reading each slide, especially when dealing with a large percentage of slides with low-density parasitemia. However, increasing slide examination time seems to be impractical under the clinic conditions described. Thus, improved quality-control measures for slide reading and blood smear staining are key to more accurate diagnosis at malaria clinics such as this one. The results of this study imply that the results of therapeutic or vaccine trials that are based on diagnoses made by field microscopists should be interpreted with caution; they underscore the need to standardize malaria smear preparation and certify every microscopist participating in important clinical trials.

Biases toward high-density detection and diagnosis pose problems on many levels, including the definition of clinical attack\textsuperscript{17} and epidemiologic classification.\textsuperscript{18} The importance of accuracy at the clinical, individual level is obvious. The effects of microscopy error on clinical trial results have been demonstrated.\textsuperscript{16} At the epidemiologic, population level, errors distort malaria statistics and mislead studies of natural immunity and the evaluation of operational interventions.

Gametocytes and mixed-species infections were problematic for AFRIMS as well as clinic microscopists, which indicates further difficulties for studies of epidemiology, entomology, and intervention effects. It is not clear why the AFRIMS readings so often disagreed about the presence of gametocytes, but the relatively low gametocyte densities (in comparison to asexual-stage densities) observed may be one of the reasons. As Muirhead-Thomson\textsuperscript{19} and Dowling and Shute\textsuperscript{14} pointed out, dramatic under-detection might explain the frequency with which non-gametocytemic patients are reported to transmit infections to \textit{Anopheles}.

The seeming deficit of mixed infections is consistent with our earlier results from Thailand and Peru\textsuperscript{11,12,22} and with studies of other areas with co-endemic \textit{P. falciparum} and \textit{P. vivax}.\textsuperscript{23} Additionally, limitations of microscopy for the detection of mixed-species infection have been well recognized.\textsuperscript{10,24} As such, results per-
taining to mixed-species infections, even by expert microscopy, must be interpreted with caution.

Received April 29, 2003. Accepted for publication July 25, 2003.

Acknowledgments: We thank the AFRIMS Department of Immunology and Medicine’s Malaria Field Team for their technical assistance, and gratefully acknowledge the contributions of G. A. Esoworthy, G. T. Keusch, D. P. Mason, and an anonymous reviewer.

Financial support: We gratefully acknowledge the support of the U.S. Army Medical Materiel Development Activity, U.S. Army Medical Research and Materiel Command (Fort Detrick, Frederick, MD).

Disclaimer: The views presented are those of the authors and do not purport to represent the views of our respective institutions.

Authors’ addresses: F. Ellis McKenzie, Room 306, Building 16, National Institutes of Health, Bethesda MD 20892, E-mail: mckenzel@mail.nih.gov. Jeeraphat Sirichaisinthop, Vector-Borne Disease Control Regional Office No. 1, Phrabuddhabat, Arasburi, Thailand. R. Scott Miller, United States Army Medical Component AFRIMS, APO AP 96546; Robert A. Gasser, Jr., Walter Reed Army Institute of Research, 503 Robert Grant Avenue, Silver Spring, MD 20910; Chansuda Wongsrichanalai, U.S. Naval Medical Research Unit 2, Kompleks P2M/PLP-LITBANGKES, Jalan Percetakan Negara No. 29, Jakarta 10570, Indonesia.

REFERENCES