Impact of Microscopy Error on Estimates of Protective Efficacy in Malaria-Prevention Trials

Colin Ohrt,[1] Purnomo,[3] M. Awalludin Sutamihardja,[3] Douglas Tang,[2] and Kevin C. Kain[4]

[1]Division of Experimental Therapeutics, and [2]Department of Biostatistics, Walter Reed Army Institute of Research, Silver Spring, Maryland; [3]United States Naval Medical Research Unit 2, Jakarta, Indonesia; [4]Toronto General Hospital and the University of Toronto, Toronto, Canada

Microscopy is an imperfect reference standard used for malaria diagnosis in clinical trials. The purpose of this study was to provide an assessment of the accuracy of basic microscopy, to compare polymerase chain reaction (PCR)–based diagnosis with microscopy results, and to assess the effect of microscopy error on apparent protective efficacy. The sensitivity and specificity of basic, compared with expert, microscopy was determined to be 91% and 71%, respectively. In a clinical trial, agreement between PCR and microscopy results improved with expert confirmation of initial results. In a simulated 12-week trial with weekly routine malaria smears, a very high specificity (>99%) for each malaria smear was found to be necessary for an estimate of protective efficacy to be within 10%–25% of the true value, but sensitivity had little effect on this estimate. Microscopy error occurs and can affect clinical trial results.

Accurate malaria diagnosis is critical in field trials evaluating antimalarial drugs or vaccines. Errors in diagnosis (false positives, false negatives, and species identification errors) may lead to biased estimates of protective efficacy. Microscopic diagnosis is considered to be the reference standard for determining the protective efficacy of prophylactic drugs or vaccines. However, microscopy is an imperfect reference standard with many inherent limitations [1], including the need for highly experienced and motivated technicians, variability in smear quality, the inability to determine malaria species at low parasitemia, and the loss of slide quality with time. Artifacts resembling malaria are common. In addition, parasitemias are often low in field trials, which leads to discrepant repeat readings. In clinical trials with periodic (i.e., weekly) screening for parasitemia, large numbers of malaria smears must be read; therefore, expert microscopists usually do not perform initial readings.

Because microscopy is an imperfect reference standard, it is difficult to estimate its true sensitivity and specificity or to evaluate new diagnostic methods [2]. Several studies have compared microscopy with new diagnostic methodologies. Although not intended to estimate the accuracy of microscopy and limited by discrepant analysis [3, 4], some estimates of the accuracy of basic and expert microscopy can be made on the basis of presented data. Reports from Thailand [5] and the Solomon Islands [6] have compared malaria basic microscopy with expert microscopy or acridine orange microscopy plus polymerase chain reaction (PCR) techniques. Data included in those articles suggest that the sensitivity of basic microscopy was low (71%–76%) and that specificity was variable (72%–95%) [5, 6]. Studies in Thailand and Kenya compared expert microscopy with repeat-smear examination and PCR [7, 8]. Sensitivity and specificity were much higher (97%–99% and 96%–99%, respectively). Four studies based on PCR techniques have revealed that mixed species infections are common (6%–19% of positive results) and often missed by both field-based and expert microscopy (50%–100% of the time) [5, 6, 8, 9].

Considerable epidemiologic and statistical literature addresses the effects of misclassification involving dichotomous variables (2×2 tables) [10–12]. The direction and magnitude of bias depends on the circumstances and the type of sampling. In randomized clinical trials, misclassification in response outcome usually leads to an underestimate of the true treatment-efficacy model [10, 11]. In the context of field trials of prophylactic drugs or vaccines, errors in diagnosing malaria do occur and can be expected to lead to biased estimates of measures of protective efficacy. The impact that these diagnostic errors and confirmation strategies may have on the outcome of malaria-

Received 4 February 2002; revised 9 April 2002; electronically published 2 August 2002.

Presented in part: 45th annual meeting of the American Society of Tropical Medicine and Hygiene, Baltimore, 1–5 December 1996 (abstract 381); 100th annual meeting of the American Society for Clinical Pharmacology and Therapeutics, San Antonio, Texas, 18–20 March 1999 (abstract PI-69); American Society of Tropical Medicine and Hygiene, Atlanta, 11–15 November 2001 (abstract 248).

Informed consent was obtained from the study subjects. The studies were approved by the United States Army, United States Navy, and Republic of Indonesia committees governing the protection of human subjects.

The views expressed here are those of the authors and not necessarily those of the United States Army or the United States Department of Defense.

Financial support: Medical Research Council of Canada (KCK MT-13721); F. Hoffmann–La Roche; United States Army Medical Research and Materiel Command; United States Naval Medical Research and Development Command.

Reprints or correspondence: Dr. Colin Ohrt, Div. of Experimental Therapeutics, Walter Reed Army Institute of Research, 503 Robert Grant Ave., Silver Spring, MD 20910-7500 (Colin.Ohrt@na.amedd.army.mil).

The Journal of Infectious Diseases 2002; 186:540–6

0022-1899/2002/18604-0013$15.00

prevention trials is largely unrecognized. In addition, limited information is published regarding procedures to improve the accuracy of microscopy in any setting. Most malaria-prevention field trials do use some method to confirm results, although the methods vary and are often not reported. The confirmation strategy used will also affect the protective efficacy reported.

As a component of a malaria prophylaxis trial, the objectives of this study were to compare the accuracy of initial microscopy with expert microscopy; to compare PCR-based diagnoses of samples collected during the trial with initial readings, expert reference readings, and final malaria diagnoses; and to assess the effects of diagnostic errors (false-positive or false-negative results) on reported estimates of protective efficacy based on a model for misclassification errors.

Materials and Methods

Data presented are based on malaria smears and blood samples obtained during a double-blind, placebo-controlled malaria prophylaxis trial in Papua (Irian Jaya), Indonesia, that was conducted between May and August 1993 [13]. Two hundred four soldiers were randomized to 3 prophylaxis arms (69 received placebo, 68 received weekly mefloquine, and 67 received daily doxycycline) for ~15 weeks. A confirmed positive malaria smear (diagnostic algorithm described below) was the primary end point for the prophylaxis trial.

Diagnosis of malaria. Giemsa-stained thick and thin malaria smears were done weekly and when the soldiers had any of the following symptoms: headache, fever, chills, nausea, or vomiting. Smears were examined by oil-immersion microscopy (magnification, ×1000) and were considered to be negative if no asexual parasites were found in 200 ocular fields of the thick film. The following procedure was used to determine the final ("confirmed" for efficacy estimates) diagnosis of malaria. Subjects with positive initial smears had a repeat malaria smear as soon as possible. If both smears were positive, then a field diagnosis of malaria was made (initial microscopy results), and the soldier was removed from the trial. Later, all positive smears and a representative sample of negative smears were read by an expert microscopist blinded to the initial result ("expert diagnosis"). Smears with discordant results were reread by the principal investigator prior to breaking the study code. A "majority rule" determined the final confirmed microscopic result. The sample of negative smears consisted of ~50 smears collected at the time a symptom consistent with malaria was reported and 250 routine weekly smears.

Microscopists. There were 4 on-site microscopists, usually 2 at the study site at a given time. Two of the 4 had been employed by the sponsoring organization as microscopists for the preceding 2 years (identified as field microscopists). The other 2 had previous experience with microscopy but were newly hired and trained (identified as new hires; table 1). The 2 field microscopists employed by the sponsoring organization read all study smears, and only their results constituted the "initial microscopy result." The reference expert microscopist (Purnomo, an author of this report) has >40 years of professional experience in the diagnosis of malaria and is internationally recognized in the field (identified as expert reference microscopist).

PCR specimen collection. Malaria culture (cryopreserve) specimens were used for PCR. Blood was obtained only from persons who had positive results for malaria smears. Molecular diagnosis and malaria species determination were done for all participants from whom blood samples were obtained. Cryopreserves were collected as whole blood in acid citrate dextrose. Packed red blood cells (1 mL) were mixed with tyrode buffer (3.4 mL) and dimethyl sulfoxide (0.6 mL) and immediately frozen in liquid nitrogen.

PCR detection and species identification. Plasmodial small-subunit rRNA genes were amplified, and species-specific oligoprobe hybridization (rDNA–oligoprobe hybridization) was done as described elsewhere [14, 15]. All PCR analyses were interpreted blinded to the microscopy results. All samples that were initially negative were reextracted and reamplified using the same method. Positive and negative blood sample controls were included with each amplification assay. To prevent cross-contamination, designated pieces of equipment and separate rooms were used for the preparation of samples and the handling of amplified products.

Statistical methods. Data were managed and tables were constructed using Microsoft Excel. The Wilcoxon rank sum test was used to compare parasite densities by use of SPSS software (version 8.0; SPSS). The effects of false-positive and false-negative microscopy errors on apparent protective efficacy were assessed by use of a modification of a model reported by Goldberg [10] and Copeland et al. [11]. Figures were constructed using Minitab 11 for Windows (Minitab). Refer to the statistical appendix for illustration of the effect of malaria misdiagnosis on the resulting estimates of protective efficacy.

Results

Accuracy of readings by newly hired microscopists. To qualify for employment as microscopists for the field trial, several individuals working as microscopists in the area undertook an examination in which 93 malaria smears (79 positive and 14 negative) were read blindly. The diagnostic errors (false positive, false negative, and species misidentification) of 2 new hires and a senior technician (not part of this project) relative to the expert reference microscopist are summarized in table 1. Individuals not hired did not perform as well as those hired (data not shown). These data illustrate that microscopy has limitations in terms of sensitivity and specificity, which vary by microscopist. Diagnostic accuracy of the new hires improved after continued intensive training (data not shown).

Table 1. Test results for newly hired and trained field microscopists, compared with results for an expert reference microscopist.

Microscopist	False positive, %	False negative, %	False species, %
Senior[a]	36	5	13
New hire 1[b]	7	18	15
New hire 2[b]	43	5	14
Mean	29	9	14

NOTE. Ninety-three blood smears were read; 79 were positive and 14 were negative.

[a] Not part of this project (for comparison only).

[b] Newly hired and trained but with previous microscopy experience.

Clinical trial sequential microscopy results. Table 2 summarizes the results of the initial, expert, and final microscopy diagnoses for the clinical trial and corresponding results based on PCR. At the time of diagnosis, 98% of subjects were symptomatic; 93% had ⩾2 symptoms typical of malaria. Parasitemias were low (median, 360 parasites/μL). *Plasmodium falciparum* parasitemias were significantly higher than those of *Plasmodium vivax* (median, 720 vs. 270 parasites/μL; $P = .02$). Ten percent of patients with *P. falciparum* and 42% of patients with *P. vivax* had ⩽120 parasites/μL.

Significant discordance in results was identified on the serial microscopy readings. On the basis of the initial reading, 55 subjects were parasitemic, 27 with *P. falciparum* and 28 with *P. vivax*. Twelve (22%) of the initial findings for the 55 subjects were discordant with the subsequent findings of the expert reference microscopist. The expert determined that 7 of the cases initially diagnosed as *P. vivax* were instead *P. falciparum* and that 4 initially diagnosed as *P. falciparum* were instead *P. vivax,* and 1 case initially read as *P. falciparum* was reread as a mixed infection. Findings for the 12 discordant smears were read again by the principal investigator before breaking the study code: 4 differed from the expert reading (2 cases diagnosed as *P. vivax* by the expert reference microscopist were reread as *P. falciparum* and 1 case read as *P. falciparum* by the expert was reread as *P. vivax*). No parasites were found in another smear that was read as very low-density *P. vivax* by the expert and as *P. falciparum* in the initial reading. Its final result was deemed to be negative. On the basis of the majority of the 3 readings, the final results were as follows: 30 *P. falciparum,* 23 *P. vivax,* 1 mixed, and 1 negative. After results were finalized and the study code was broken, 3 additional expert microscopists were asked to examine the smear deemed negative by the principal investigator (a smear from a doxycycline recipient) and the smear from the single doxycycline prophylaxis failure. No malaria parasites were found in either case.

Microscopy compared with PCR. Blood samples for PCR were obtained only at the time of the initial diagnosis of malaria. Samples were available for 53 of the 55 subjects. All 53 were blindly assessed for a species-specific diagnosis by PCR.

PCR results compared with the sequential microscopy readings are presented in table 2. On the basis of the initial microscopy results, 27 cases of *P. falciparum* and 28 cases of *P. vivax* were identified. Of the 27 *P. falciparum* cases, 26 were available for PCR analysis. Of these, complete concordance (*P. falciparum* only) was shown in 20 (77%), partial concordance (mixed *P. falciparum* and *P. vivax*) was shown in 2 (8%), and discordance was shown in 4 (15%) cases. Of the discordant isolates, 2 were *P. vivax* and 2 were negative by PCR. As a comparison, of the 26 cases determined to be *P. falciparum* by PCR, 20 (77%) were diagnosed as *P. falciparum* by microscopy. Of the 28 cases of *P. vivax,* 27 were available for PCR analysis. Of these, 18 (67%) were completely concordant, 3 (11%) were partially concordant, and 6 (22%) were discordant. All the discordant results were *P. falciparum*. Of the 20 cases of *P. vivax* determined by PCR, 18 (90%) were diagnosed as *P. vivax* by microscopy. Of 5 mixed infections determined by PCR, none was identified by microscopy.

On the basis of expert reference microscopy, 28 cases of *P. falciparum* were identified (table 2). Of these 28, there was complete concordance (*P. falciparum* only) in 23 (82%), partial concordance (mixed *P. falciparum* and *P. vivax*) in 3 (11%), and discordance in 2 (7%). Of the discordant isolates, 1 was *P. vivax* and 1 was negative by PCR. Of the 23 *P. vivax* diagnoses by the expert reference microscopist, 19 (83%) were completely concordant, 2 (9%) were partially concordant, and 2 (9%) were discordant. One discordant result was *P. falciparum,* and the other was negative. On the other hand, of the 25 cases of *P. falciparum* determined by PCR, 23 (92%) were diagnosed as *P. falciparum* by expert reference microscopy. Of the 20 cases of *P. vivax* determined by PCR, 19 (95%) were diagnosed as *P. vivax* by microscopy. Of the 5 mixed infections by PCR, none was identified by microscopy.

The 12 discordant readings between the initial microscopic reading and the expert reference reading results were reread a third time by the principal investigator. Four results (33%) by the principal investigator differed from those of the expert. PCR results were the same as the expert's reading in 2 of the 4 cases. In the third case, the initial microscopic reading was *P. falciparum,*

Table 2. Results of the initial, expert, and final microscopy diagnoses in a malaria-prevention clinical trial and corresponding results based on polymerase chain reaction (PCR).

	Microscopy results										
	Initial diagnosis			Expert reference diagnosis			Final diagnosis				
PCR diagnosis	*P. falciparum*	*P. vivax*	Mixed	*P. falciparum*	*P. vivax*	Mixed	*P. falciparum*	*P. vivax*	Mixed	Negative	Total
P. falciparum	20	6	0	23[a]	1	1	24	1	1	0	26
P. vivax	2	18[b]	0	1	19[b]	0	1	19[b]	0	0	20[b]
Mixed	2	3	0	3	2	0	3	2	0	0	5
Negative	2	0	0	1	1	0	1	0	0	1	2
No PCR sample	1	1	0	1	1	0	1	1	0	0	2
Total	27	28	0	29	24	1	30	23	1	1	55
Agreement with PCR results, %	77	67	0	82	83	0	83	86	0	100	

NOTE. *P. falciparum, Plasmodium falciparum; P. vivax, Plasmodium vivax*.

[a] One slide pair was lost and not read by the expert; the final diagnosis assumed that the original reading was correct.

[b] Includes 1 sample with a PCR product but no specific diagnosis.

the expert reference reading was *P. vivax,* and the principal investigator's reading and PCR were both negative. In the last case, the expert reference reading was *P. vivax,* whereas the other 2 readings were *P. falciparum.* PCR revealed *P. falciparum.*

Effect of diagnostic strategy on apparent protective efficacy. When routine smears are collected from asymptomatic patients, false-positive results can have a profound effect on protective efficacy (figure 1*A*). Methods to improve specificity may be at the cost of sensitivity. However, decreased sensitivity (false-negative results) does not significantly impact the estimate of protective efficacy (figure 1*B*). The statistical appendix details the model used to assess the impact of false-positive and false-negative results illustrated in figure 1.

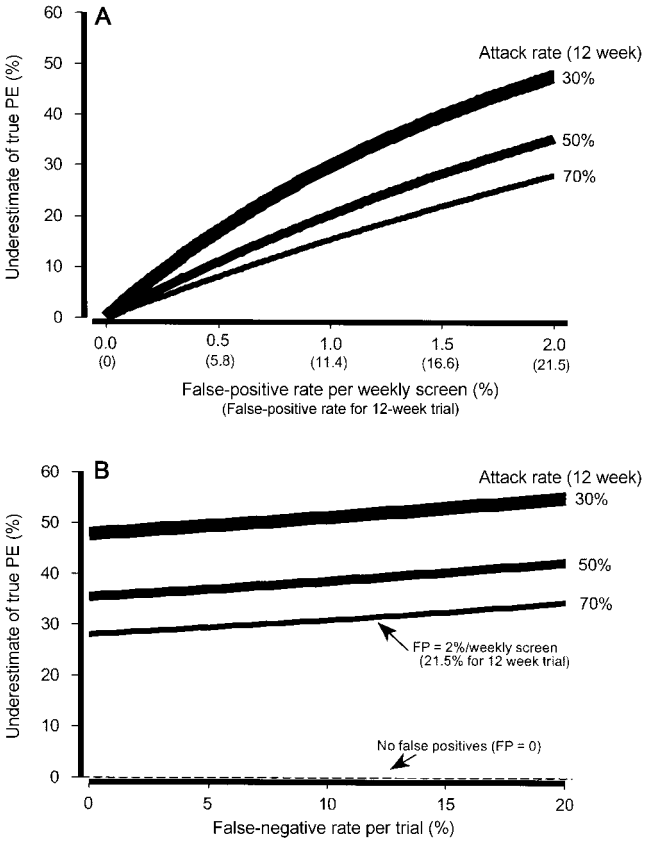

Figure 1. *A,* The underestimate of the true protective efficacy (PE) caused by false-positive (FP) malaria smears in a 12-week malaria-prevention trial with routine weekly collection. Note that 0.5% FP smears (99.5% specificity) cause the PE to be underestimated by 7%–15%. As shown, the underestimate varies with the malaria attack rate. *B,* The underestimate of the true PE caused by false-negative malaria smears in a 12-week malaria-prevention trial with routine weekly collection. Note that false-negative smears have no effect on PE estimates when no false-positive results occur. A slight effect is seen with false-negative results at the varying malaria attack rates when false-positive results are present.

Discussion

To our knowledge, this is the first study critically assessing microscopy in the context of a malaria-prevention trial. Our results (table 1), experience, and review of the literature strongly suggest that false-positive results occur with microscopy [5–8] and that diagnostic errors persist in the clinical trial setting. In addition, this is the first assessment of how the sensitivity and specificity of microscopy-based diagnosis may impact the reported efficacy results of such trials (figure 1). We have identified that the lower-than-expected protective efficacy found in some clinical trials was indeed due to the predictable underestimation of efficacy resulting from false-positive malaria smears.

Newly hired microscopists had substantial error on a microscopy examination (table 1). Many artifacts closely resemble malaria parasites, which may have led to the false-positive readings. However, because the "negative" test smears were obtained in an area endemic for malaria, it is not definitely known that these are truly negative. In the clinical trial presented in this manuscript [13], one false-positive smear finding occurred in the initial, compared with the final, microcopy diagnosis, and 2 occurred by PCR analysis. About 2284 smears were negative by initial microscopy. If one assumes that these were negative by PCR, specificity in the trial was ~99.9%. Overall sensitivity for this trial was also likely very high. Unlike some clinical trial settings, false-negative results were likely to be detected when subjects had additional malaria smears for symptoms of malaria.

Compared with both the expert reference microscopy and PCR, the most common error in initial readings was species identification errors. Most commonly, *P. falciparum* was misidentified as *P. vivax*. We believe that many of these errors were due to less-experienced field readers, specific diagnosis often based on the thick smear, and variability in smear preparation and stain quality. Rereading of all positive field diagnoses by an expert reference microscopist led to a substantial improvement in the final diagnosis. However, missed mixed infections and species errors still occurred, compared with analyses by PCR. These errors are likely inherent limitations of light microscopy for malaria diagnosis. However, with the desired end point of both *P. falciparum* and *P. vivax* prevention in the mefloquine versus doxycycline trial, the use of PCR confirmation would not have significantly impacted results.

In this field trial, at least one field microscopist was always at the study location. However, a quality assurance program was not in place at the time of this study and likely contributed to the errors in the initial reading. Good clinical practice [16] is essential for malaria diagnosis in clinical trials. This should include routine, on-going training and recertification of technicians through testing with slides collected in the field (negative smears should be collected in a malaria-free area). Standard operating procedures should be written for every aspect of slide preparation, quality assessment, reading, recording of results, and the storage of slides. Slides of unacceptable quality should be rejected and repeated. Expert confirmation of results is nec-

essary and must be completely blinded to avoid the introduction of bias. In addition, because of deterioration of slide quality with time, photographic or digital recording of results may be the preferred method of permanent documentation and possibly an end point. The use of PCR and rapid diagnostic tests, such as dipsticks, in defining clinical trial end points or in confirming microscopy results should be studied. Dipsticks have the advantage of simple, immediate diagnosis or confirmation of results. However, the sensitivity and specificity of the currently available devices appear to be lower than that of microscopy- and PCR-based methods [15, 17].

In this double-blind, placebo-controlled trial, all microscopy diagnoses were finalized before the study code was broken. All PCR samples were also interpreted blindly. One limitation of this study was that PCR samples were collected as malaria cryopreserves. A second limitation is that no negative controls were collected from field isolates. Future trials should collect whole blood for PCR from a sample of subjects without malaria, perhaps subjects presenting with other febrile illnesses.

As illustrated in this study, diagnostic errors occur with microscopy, and errors can have a considerable impact in underestimating the protective efficacy of prophylactic drugs or vaccines. The sensitivity, specificity, and species identification error rate for each study technician should be assessed by certification examination and should be monitored throughout the trial. On the basis of these sensitivity and specificity estimates, the possible underestimation in the resulting protective efficacy or other effect measures due to these diagnostic errors can be assessed. Strategies that maximize final specificity are essential for malaria-prevention studies (e.g., rereading paradigm). Specific strategies will depend on the study end points. Finally, the diagnostic procedures for determining and confirming final study end points should always be reported.

Acknowledgments

We are grateful for the help of many officials of the Indonesian Army, Provincial Health Service, and Ministry of Health. Special thanks are extended to the commanders and men of battalions 143 and 731 for their support and cooperation. We also thank Slamet Harjosuwarno and Budi Subianto in Jayapura and Suriadi Gunawan; Harijani Marwoto, Sri Oemijati, and P. R. Arbani in Jakarta; and Laura Mirabelli-Primdahl, Kathleen Zhong, and Suradi, for technical assistance.

Appendix

Statistical Appendix

Several investigators have considered the effect of misclassification errors (in 2×2 tables) on effect measures (e.g., odds ratios, relative risk, and risk difference) in epidemiologic investigations [10–12]. However, the impact of these errors on estimates of measures of prophylactic efficacy (PE) has not received attention in the design of field trials of antimalarial drugs or in the analysis and reporting of trial results. In this appendix, we numerically illustrate the effect of malaria misdiagnosis on the resulting estimate of PE based on the percentage reduction in cumulative risk and provide the equations used to estimate the percentage underestimation of the true PE (as illustrated in figure 1) that can occur in the presence of errors in the final diagnosis of malaria. These include the cumulative effect of periodic (e.g., weekly) screening on the total false-positive rate.

Example 1. Consider a 12-week field trial comparing an antimalarial drug ($n = 100$) to placebo ($n = 100$). Assume (1) that the cumulative risk or attack rate (AR) of malaria in the placebo arm is 50%, (2) that the "true" PE of the drug is 80% (drug AR, 10%), (3) a 10% cumulative false-positive rate (0.9% per week) for both drug and placebo, and (4) no false-negative results. Under these assumptions, table A1 summarizes the (expected) results by true malaria status, diagnosis, and drug arm.

Had there been no false-positive results (9 in the drug arm and 5 in the placebo arm), the "expected" ARs in the drug and placebo arms (last row totals in table A1) would be (correctly) 10% (10/100) and 50% (50/100), leading to the correct estimate of the true PE (80%). However, because of diagnostic errors, the observed ARs (based on the malaria diagnosis) in the drug and placebo groups would be expected to be 19% (19/100) and 55% (55/100), respectively (totals for drug and placebo arms in table A1). These observed results (with 14 false-positive results) would lead to an estimate of the drug's PE of 65.5% (1–0.19/0.55), which underestimates the true PE by 18.1%.

Example 2. If, in addition to the assumptions of example 1, we also assume a false-negative rate of 10% (sensitivity, 90%) the expected outcomes would be as in table A2. In the presence of both false-positive and false-negative errors (both 10%), the estimate of the PE based on the malaria diagnosis data (malaria and no malaria totals in table A2) would be 64% (1–0.18/0.50), which underestimates the true PE by 20%. Note that the equal false-positive and false-negative errors are not offsetting, and the effect of false-negative errors had little impact over the effect caused by the presence of false positives.

Table A1. Summary of expected results by true malaria status, diagnosis, and drug arm in a 12-week field trial comparing an antimalarial drug with placebo.

	True malaria status					
	Drug ($n = 100$)			Placebo ($n = 100$)		
Diagnosis	Malaria	No malaria	Total	Malaria	No malaria	Total
Malaria	10	9[a]	19	50	5[a]	55
No malaria	0	81	81	0	45	45
Total	10	90	100	50	50	100

NOTE. Results assume that the cumulative risk or attack rate (AR) of malaria in the placebo arm is 50%, that the "true" prophylactic efficacy of the drug is 80%, a 10% cumulative false-positive rate for both drug and placebo, and no false-negative results.

[a] False positive.

Table A2. Summary of expected results by true malaria status, diagnosis, and drug arm in a 12-week field trial comparing an antimalarial drug with placebo.

	True malaria status					
	Drug (n = 100)			Placebo (n = 100)		
Diagnosis	Malaria	No malaria	Total	Malaria	No malaria	Total
Malaria	9	9[a]	18	45	5[a]	50
No malaria	1[b]	81	82	5[b]	45	50
Total	10	90	100	50	50	100

NOTE. Results assume that the cumulative risk or attack rate (AR) of malaria in the placebo arm is 50%, that the "true" prophylactic efficacy of the drug is 80%, a 10% cumulative false-positive and false-negative rate for both drug and placebo.
[a] False positive.
[b] False negative.

The above examples illustrate that the presence of diagnostic errors (especially false-positive errors) can lead to seriously biased estimates of a drug's PE (in this case, the percentage reduction in cumulative risk over the period of the trial). In general, the magnitude (and direction) of the bias (absolute or relative) depends on the sensitivity and specificity of the diagnostic procedure, the baseline (placebo) cumulative AR, and whether the diagnostic errors are the same or different (nondifferential or differential) in the drug and placebo arms. In the case of nondifferential diagnostic errors (same for drug and placebo), the result will be to underestimate the true PE.

Statistical model for effects of diagnostic errors on estimated PE. Table A3 symbolically represents the outcomes from a randomized placebo-controlled trial of an antimalarial drug. On the basis of the observed numbers of diagnosed (apparent) malaria cases in the placebo and drug arms (c_0 and c_1, respectively) and corresponding sample sizes (n_0 and n_1, respectively), the estimated cumulative risks (ARs) in the drug and placebo arms are (c_1/n_1) and (c_0/n_0), respectively. The usual "estimate" of the PE, based on the percentage reduction in estimated cumulative risk, is $\%\ \mathrm{PE} = [1 - (c_1/n_1)/(c_0/n_0)]100$. Note that this estimate assumes no loss to follow-up.

To find the expected bias in the above estimate of PE under the assumption of nondifferential (assumed same in placebo and drug) diagnostic errors, let

$$\Pi_0 = \text{true placebo cumulative malaria risk},$$

$$\Pi_1 = \text{true drug cumulative malaria risk},$$

$$\mathrm{PE}_{\mathrm{true}} = \text{true PE} = 1 - \Pi_1/\Pi_0,$$

$$\mathrm{PE}_{\mathrm{app}} = \text{expected (apparent) value of PE} = \mathrm{E(PE)},$$

$$\alpha = \text{probability of false positive},$$

$$\beta = \text{probability of false negative}.$$

When the notation above is used, $\mathrm{PE}_{\mathrm{app}} = \mathrm{E(PE)} = \mathrm{E}[1 - (c_1/n_1)/(c_0/n_0)]$.

To a first order of approximation (propagation of error), the expected value of PE ($=\mathrm{PE}_{\mathrm{app}}$) is obtained by replacing the proportions of diagnosed cases (c_0/n_0 or c_1/n_1) with their corresponding expected values:

$$\mathrm{E}(c_0/n_0) = \Pi_0(1-\beta) + (1-\Pi_0)\alpha$$

and

$$\mathrm{E}(c_1/n_1) = \Pi_0(1-\mathrm{PE}_{\mathrm{true}})(1-\beta) + [1-\Pi_0(1-\mathrm{PE}_{\mathrm{true}})]\alpha\ .$$

After some algebra, we obtain the following relationship between the apparent (expected) PE (in presence of diagnostic errors) and the true PE (in absence of diagnostic errors):

$$\mathrm{PE}_{\mathrm{app}} = \mathrm{E(PE)} = \mathrm{PE}_{\mathrm{true}}\left[\frac{\Pi_0(1-\alpha-\beta)}{\alpha + \Pi_0(1-\alpha-\beta)}\right]\ . \qquad (1)$$

Equation (1) is based on a similar model for bias used by Copeland [11] to estimate the underlying (true) relative risk of disease occurrence in cohort studies using estimated sensitivity and specificity of the classification procedure.

Two immediate consequences can be seen from equation (1) above. First, if as is likely to be the case, the false-positive or false-negative rates are such that $\alpha + \beta < 1$, then the expression in square brackets is always <1, and, hence, the usual estimate of PE underestimates the true efficacy ($\mathrm{PE}_{\mathrm{app}} < \mathrm{PE}_{\mathrm{true}}$). Second, in the absence of false-positive errors, $\alpha = 0$ (100% specificity), the usual estimate is unbiased ($\mathrm{PE}_{\mathrm{app}} = \mathrm{PE}_{\mathrm{true}}$). This clearly indicates the importance of minimizing the probability of a false positive (maximizing diagnostic specificity). In the case of nondifferential diagnostic errors the percent bias (under estimation of $\mathrm{PE}_{\mathrm{true}}$) is given by equation (2):

$$\%\ \text{Bias} = \frac{\mathrm{PE}_{\mathrm{true}} - \mathrm{PE}_{\mathrm{app}}}{\mathrm{PE}_{\mathrm{true}}} = \left\{1 - \left[\frac{\Pi_0(1-\alpha-\beta)}{\alpha + \Pi_0(1-\alpha-\beta)}\right]\right\} 100$$

$$= \left[\frac{\alpha}{\alpha + \Pi_0(1-\alpha-\beta)}\right] 100\ . \qquad (2)$$

Effects of periodic screening on the overall false-positive rate. In equations (1) and (2), the false-positive (α) and false-negative (β) rates correspond with the overall (cumulative)

Table A3. Symbolic representation of the outcomes from a randomized placebo-controlled trial of an antimalarial drug.

Diagnosis	Drug	Placebo
Malaria	c_1	c_0
No malaria	$n_1 - c_1$	$n_0 - c_0$
Total	n_1	n_0

NOTE. c_0 and c_1, no. of diagnosed (apparent) malaria cases in the placebo and drug arms, respectively; n_0 and n_1, no. of persons in the placebo and drug arms, respectively.

probability of misdiagnosis during the follow-up period of the trial. In most drug trials, follow-up consists of periodic (e.g., weekly) routine screens for parasitemia. Even if the false-positive rate at each screen is assumed to be acceptably low, the effects of multiple testing of persons who do not get infected during the period of the trial (true negatives) can be unacceptably high. If "α_s" is the false-positive rate at each screen and there are k periodic screens, then (assuming independence) the total false-positive rate (α) will be $\alpha = 1 - (1 - \alpha_s)^k$.

For example, in a 12-week trial with weekly screens ($k = 12$) and a per screen false-positive rate of 1% ($\alpha_s = 0.01$), the overall false-positive rate (α) is $1 - (1 - .01)^{12} = 0.114$ or 11.4%. There is no simple way to quantify the effects of multiple testing on the total false-negative rate (β).

References

1. Payne D. Use and limitations of light microscopy for diagnosing malaria at the primary health care level. Bull World Health Organ **1988**; 66:621–6.
2. Valenstein PN. Evaluating diagnostic tests with imperfect standards. Am J Clin Path **1990**; 93:252–8.
3. Miller WC. Bias in discrepant analysis: when two wrongs don't make a right. J Clin Epidemiol **1998**; 51:219–31.
4. Hadgu A. The discrepancy in discrepant analysis. Lancet **1996**; 348:592–3.
5. Snounou G, Viriyakosol S, Jarra W, Thaithong S, Brown KN. Identification of the four human malaria parasite species in field samples by the polymerase chain reaction and detection of a high prevalence of mixed infections. Mol Biochem Parasitol **1993**; 58:283–92.
6. Arai M, Kunisada K, Kim H, et al. A colorimetric DNA diagnostic method for falciparum malaria and vivax malaria: a field trial in the Solomon Islands. Nucleosides Nucleotides **1996**; 15:719–31.
7. Barker RH Jr, Banchongaksorn T, Courval JM, Suwonkerd W, Rimwungtragoon K, Wirth DF. A simple method to detect *Plasmodium falciparum* directly from blood samples using the polymerase chain reaction. Am J Trop Med Hyg **1992**; 46:416–26.
8. Oliveira DA, Holloway BP, Durigon EL, Collins WE, Lal AA. Polymerase chain reaction and a liquid-phase, nonisotopic hybridization for species-specific and sensitive detection of malaria infection. Am J Trop Med Hyg **1995**; 52:139–44.
9. Brown AE, Kain KC, Pipithkul J, Webster HK. Demonstration by the polymerase chain reaction of mixed *Plasmodium falciparum* and *P. vivax* infections undetected by conventional microscopy. Trans R Soc Trop Med Hyg **1992**; 86:609–12.
10. Goldberg JD. The effects of misclassification on the bias in the difference between two proportions and the relative odds in the fourfold table. J Am Stat Assoc **1975**; 70:561–7.
11. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. Am J Epidemiol **1977**; 105: 488–95.
12. Fleiss JL, Shrout PE. The effects of measurement errors on some multivariate procedures. Am J Public Health **1977**; 67:1188–91.
13. Ohrt C, Richie TL, Widjaja H, et al. A double-blind placebo-controlled trial of mefloquine versus doxycycline for the prophylaxis of malaria in Indonesian soldiers. Ann Intern Med **1997**; 126:963–72.
14. Li J, Wirtz RA, McConkey GA, et al. *Plasmodium:* genus-conserved primers for species identification and quantitation. Exp Parasitol **1995**; 81:182–90.
15. Humar A, Ohrt C, Harrington MA, Pillai D, Kain KC. Parasight F test compared with the polymerase chain reaction and microscopy for the diagnosis of *Plasmodium falciparum* malaria in travelers. Am J Trop Med Hyg **1997**; 56:44–8.
16. International Conference on Harmonization. Good clinical practice: consolidated guideline. Washington, DC: Federal Register 9 May **1997**. FR Document No. 25691.
17. Pieroni P, Mills CD, Ohrt C, Harrington MA, Kain KC. Comparison of the ParaSight-F test and the ICT Malaria Pf test with the polymerase chain reaction for the diagnosis of *Plasmodium falciparum* malaria in travellers. Trans R Soc Trop Med Hyg **1998**; 92:166–9.