TRANSACTIONS OF THE ROYAL SOCIETY OF TROPICAL MEDICINE AND HYGIENE (2002) 96, 254–257

Validity, reliability and ease of use in the field of five rapid tests for the diagnosis of *Plasmodium falciparum* malaria in Uganda

J. P. Guthmann[1]*, A. Ruiz[2], G. Priotto[1], James Kiguli[3], L. Bonte[2] and D. Legros[1] *[1]Epicentre, 4 rue Saint Sabin, 75011 Paris, France; [2]Médecins sans Frontières, 8 rue Saint Sabin, 75011 Paris, France; [3]Mbarara University of Science and Technology, Mbarara, Uganda*

Abstract

A study was conducted to measure the overall performance of several rapid diagnostic tests for *Plasmodium falciparum* infection, in order to select the most appropriate test to be used in the field. A total of 742 patients attending the out-patient department of Mbarara Hospital with a clinical suspicion of malaria were included in the study. For each patient, a thick/thin film and 5 rapid tests based on the detection of histidine-rich protein II (HRP-II) (Paracheck Pf dipstick and device, ParaHIT f, Malaria Rapid and BIO P.F.) were performed. Outcomes were validity, inter-reader reliability and 'ease of use in the field', measured by both the general characteristics of the test and by the opinion of the readers. About half (57%) of the patients were positive for *P. falciparum*. The Paracheck Pf (dipstick and device) was considered as the most appropriate for the use in the field, being sensitive (97%), moderately specific (88%), reliable (kappa coefficient = 0·97), easy to use and cheap (about US$ 0.5/test). The ParaHIT f represented a good alternative.

Keywords: malaria, *Plasmodium falciparum*, diagnosis, rapid tests, comparisons, sensitivity, specificity, positive predictive value, negative predictive value, Uganda

Introduction

Rapid diagnostic tests (RDTs), based on the detection of antigens derived from malaria parasites, represent a good alternative to microscopy for the diagnosis of *Plasmodium falciparum* malaria (WHO, 1996). RDTs are reasonably sensitive and specific, and are easy to perform and to interpret. Most of the studies, however, have focused on measuring only validity (sensitivity, specificity and predictive values). Very few studies have evaluated reliability of a test (KILIAN *et al.*, 1999; LEMA *et al.*, 1999), and no studies have balanced validity and reliability with other practical characteristics for ease of use in the field. This is important since a test may be valid but expensive, or valid in experimental conditions but with drawbacks making it unsuitable for its use in the field.

In light of these omissions, we conducted a study to compare several tests available, with the objective of selecting the most appropriate one to be used in the field if it is valid, reliable, easy to use and cheap. This evaluation was done in usual field conditions, in a large sample of patients, clinically suspected to have malaria, attending an out-patient department facility in Uganda.

Methods

Study area

The study was conducted in Mbarara, a city of 52 026 inhabitants located 300 km southwest of Kampala (Uganda). Patients were recruited at the out-patients department of Mbarara Regional Hospital, which is the reference hospital for the south-western region of Uganda. Diagnosis of malaria is mainly clinical.

Study patients

Every patient presenting at the out-patient department with a clinical suspicion of malaria (as diagnosed by the clinical officer) was included in the study. There were no exclusion criteria, other than refusal to participate. Demographic and clinical information was recorded, and thick/thin blood smear and RDTs were performed. Every person positive for malaria (as assessed by the parasitological examination) was treated with a combination of artesunate and sulfadoxine–pyrimethamine.

Sample size

The required number of blood smear-positive patients was calculated according to the following parameters: expected sensitivity of a test = 90%, precision = 5%, alpha error = 0·05. This number (n = 136) was doubled taking into account a stratified analysis by age-group category (0–4, ≥5 years). Similar parameters (with expected specificity of 90%) were used to calculate the required number of blood smear-negative patients. The minimal final sample size was therefore fixed at 300 positive and 300 negative patients, i.e., 600 individuals.

Blood and urinary tests

The blood smear results were considered as the 'gold standard' against which the results of the RDTs were compared. Blood samples were collected by fingerprick from clinically suspected malaria cases. Thin and thick blood films were stained with 2% Giemsa and analysed microscopically by a trained malaria technician. Parasitaemia was counted against 200–500 white blood cells, and a slide was considered negative after looking at 200 high-power fields. Results were blind. Results of the blood smear were quantified in parasites/μL, and considered either (a) positive for *P. falciparum* by the presence of asexual forms of *P. falciparum*, or (b) negative for *P. falciparum* in the absence of asexual parasites of *P. falciparum* regardless of the presence of other parasite species or *P. falciparum* gametocytes. A quality-control procedure was put in place: all discordant results (between the RDT and the slide), all slides where only *P. falciparum* gametocytes were detected and a random sample of 20% of the remaining slides were checked blind by an independent trained laboratory technician.

Five RDTs to be evaluated were selected on the basis of 2 main criteria: (a) tests detecting *P. falciparum* antigens alone; (b) tests costing about US$ 1 or less. All the tests selected detect *P. falciparum* histidine-rich protein II (HRP-II) in blood: BIO P.F. (VEDA LAB, France), Malaria Rapid (Vision Biotech, South Africa), Paracheck Pf dipstick and device (Orchid Biomedical Systems, India) and ParaHIT f (Span Diagnostics Ltd, India). Tests were performed according to the manufacturers' instructions, and read twice after 15–30 min by a trained nurse and by the research assistant. Readers were kept blinded to the result of the microscopy examination (which was given 30–60 min later) and to the other reader's verdict. A test was considered as invalid when the control line was missing and doubtful when the pink line was faint or incomplete. At the end of the study, the 2 readers gave their opinion about each test, concerning the ease of use and interpretation.

*Author for correspondence; fax +33 1 40212803, e-mail jguthmann@epicentre.msf.org

A Saker–Solomons test for the detection of anti-malarial drugs (principally chloroquine and its metabolites) (MOUNT *et al.*, 1989) was performed in individuals producing a urine sample. A yellow–green colour indicated a negative test, a red to purple colour indicated a positive result.

Outcomes under investigation

Sensitivity was defined as the number of individuals with a positive test among the total number of individuals with a positive blood slide. Specificity was defined as the number of individuals with a negative test among the total number of individuals with a negative blood slide. Positive predictive value (PPV) was defined as the number of individuals with a positive blood slide among the total number of individuals with a positive test. Negative predictive value (NPV) was defined as the number of individuals with a negative blood slide among the total number of individuals with a negative test. For these calculations, only the result of the first reading of the RDT was taken into account (the second reading was done only to assess reliability). Inter-reader reliability measured the extent to which the reading of the test differed from one observer to the other. A series of characteristics that were considered important for the use of an RDT in the field were listed before the beginning and recorded at the end of the study. These were: (a) general characteristics of the test (number of invalid and of doubtful tests, shelf-life, storage temperature, storage volume, cost, required blood quantity, number of steps needed for its performance, additional material needed, and stability of the result); and (b) characteristics related to the opinion of the readers (rapidity and ease of performance, ease of box and bag opening, and quality of the instruction sheet). The opinion of the 2 readers was sought independently through a structured questionnaire.

Analysis

Clinical and demographic data were recorded on an individual Patient Record Form. Results of the RDT and of the microscopy examination were recorded on separate sheets. After data entry, the database was checked against the source documents, and errors were corrected. Data were then analysed using EpiInfo software (CDC, USA; WHO).

The characteristics of the sample at entry were summarized in a Table, and compared between age-groups (0–4, ≥5 years) using statistical tests (either χ^2 test for the comparison of proportions, or ANOVA for the comparison of means). Validity results were expressed by age-group and by parasite density categories, indicators were expressed with their 95% confidence interval (95% CI) and comparisons were done using the χ^2 test. Reliability was expressed by the kappa coefficient, which was calculated for each test. A test was considered as reliable if kappa was ≥0·8. A score summarized the characteristics of a test concerning its use in the field. This score was calculated by adding specific scores given to each characteristic. For example the required quantity of blood needed was scored either 2 (if 5 μL) or 1 (if 25 μL); shelf-life was scored either 2 (>12 months) or 1 (≤12 months); etc. Adding the specific scores given to each characteristic made the final score for a test. The tests easier to use were the ones with the higher score.

A written consent to be signed by each participant (or parent/guardian) was translated into the local language before the implementation of the study. Protocol and consent were approved by the Ugandan National Council for Science and Technology (Ethics Committee).

Results

Between 28 November 2000 and 29 January 2001, 742 patients were recruited into the study, of whom 315 (42%) were aged <5 years. About half the patients (423, 57%) had a positive blood smear for asexual forms of *P. falciparum* (385 *P. falciparum* monoinfections, 2 mixed infections *P. falciparum* + *P. vivax*, and 36 mixed infections *P. falciparum* + *P. malariae*). The 319 negative slides for asexual forms of *P. falciparum* included 16 slides with *P. falciparum* gametocytes alone, 4 slides with *P. malariae* monoinfection and 299 slides with no parasites detectable.

The 2 Paracheck Pf and ParaHIT f had the same sensitivity (97%, 95% CI 95–98) and specificity (88%, 95% CI 83–91) (Table 1). The 2 other tests had either a low sensitivity (BIO P.F.), or a low specificity (Malaria Rapid). Overall, the sensitivities were significantly lower for parasitaemias ≤100 parasites/μL compared to parasitaemias >100 parasites/μL ($P < 0{\cdot}01$ for each test) (Table 2). In each category of parasite density, sensitivity did not change significantly between age-groups (except for the BIO P.F. for which the sensitivity was significantly higher in the 0–4-years age-group). Specificity of each test was not significantly different in the 0–4-years age-group compared to the ≥5-years age-group: 88·5% versus 87·9% ($P = 0{\cdot}87$) for the Paracheck Pf dipstick, 88·4% versus 87·3% ($P = 0{\cdot}78$) for the Paracheck Pf device, 98·3% versus 86·2% ($P = 0{\cdot}43$) for the ParaHIT f, 91·2% versus 94·2% ($P = 0{\cdot}31$) for the BIO P.F., and 79·6% versus 73·1% ($P = 0{\cdot}21$) for the Malaria Rapid.

The number of disagreements was less than 5% for all the 5 tests (Table 1). This number, however, was 2 times higher for the BIO P.F. and 3 times higher for the Malaria Rapid, compared to the 2 Paracheck Pf and the ParaHIT f. For the 5 tests, the kappa coefficient was at least equal to 0·90.

The 2 Paracheck Pf and the Malaria Rapid had practical advantages that made them more suitable for the use in the field compared to the 2 other tests (Table 3). The Paracheck Pf device seemed more difficult to perform than the dipstick, especially with babies because of the need for taking the blood sample before performing the test (whereas with the dipstick, the contact between test and finger is made directly). Main disadvantages of the BIO P.F. were a short shelf-life, a large storage volume required, a large quantity of blood required (which is inconvenient with small babies since it becomes more difficult to obtain the blood sample), and a high price. The ParaHIT f had a shorter shelf-life than the Paracheck Pf or the Malaria Rapid; it also required a large volume for its storage and, additionally, was conditioned in a bag which was difficult to open, especially wearing gloves.

Discussion

From the 5 tests evaluated, the Paracheck Pf (dipstick and device) and ParaHIT f were considered most appropriate to use in the field, being highly reliable and very sensitive for the diagnosis of *P. falciparum* malaria (especially for levels of parasitaemias above 100 parasites/μL). On the other hand, their specificity was much lower than their sensitivity. Having a relatively low specificity (which leads to an over-diagnosis and to an over-treatment of non-malaria cases) was, however, considered as less serious than having a low sensitivity (which may lead to a potentially fatal condition being missed). This relatively low specificity was sometimes explained by the presence of *P. falciparum* gametocytes [i.e., 12 out of 38 false-positive (31·6%) Paracheck Pf dipstick results had gametocytes alone in the blood, versus an overall proportion of 43 (20·2%) of 213]. It may also be explained by the persistence of HRP-II in patients who had been treated for malaria in the 2 previous weeks, as suggested by a decreased specificity in patients with a positive Saker–Solomons test, although not statistically significant. Finally, a low specificity can also be found with a very sensitive test, i.e., able to detect malaria antigens in the presence of

Table 1. Validity and inter-reader reliability of 5 rapid diagnostic tests for the detection of *Plasmodium falciparum* infection in patients attending Mbarara out-patient department (Mbarara Hospital), south-western Uganda (November 2000–January 2001)

	Paracheck Pf dipstick ($n = 741$)	Paracheck Pf device ($n = 741$)	ParaHIT f ($n = 738$)	BIO P. F. ($n = 739$)	Malaria Rapid ($n = 721$)
Validity					
Sensitivity	97·4 (95·2–98·6)	97·2 (95·0–98·4)	97·6 (95·6–98·8)	89·5 (86·1–92·2)	98·3 (96·5–99·3)
Specificity	88·1 (83·9–91·3)	87·7 (83·4–91·0)	87·3 (83·0–90·6)	93·1 (89·6–95·5)	75·3 (70·0–80·0)
Reliability					
Number of disagreements					
Positive slides	6	4	4	20	5
Negative slides	4	6	8	3	29
Total	10	10	12	23	34
Kappa coefficient	0·97	0·97	0·97	0·94	0·90

n indicates the number of interpretable tests. Other values in parentheses are 95% confidence intervals.

Table 2. Sensitivity of 5 rapid diagnostic tests for the detection of *Plasmodium falciparum* infection according to age-group and parasite density in patients attending Mbarara out-patient department (Mbarara Hospital), south-western Uganda (November 2000–January 2001)

	≤100 parasites/μL				>100 parasites/μL				
Test	Overall	0–4 years	≥5 years	*P* value	Overall	0–4 years	≥5 years	*P* value	*P* value[a]
Paracheck dipstick	86·3	95·2	80·0	0·25	98·9	100	97·9	0·14	<0·01
Paracheck device	88·2	95·2	83·3	0·40	98·4	98·9	97·9	0·73	<0·01
ParaHIT f	88·2	95·2	83·3	0·40	98·9	100	97·9	0·14	<0·01
BIO P.F.	62·7	85·7	46·7	<0·01	93·2	96·6	90·0	0·01	<0·01
Malaria Rapid	90·2	95·2	86·7	0·60	99·5	100	98·9	1·00	<0·01

[a]For each test, comparison of overall sensitivity in each category of parasite density.

very low parasite densities, undetectable by microscopy (which may be an explanation for the high sensitivity/low specificity found for the Malaria Rapid). Of the 2 Paracheck Pf and ParaHIT f, equally valid and reliable, the 2 Paracheck Pf had several characteristics that made them easier to use in the field, and therefore were considered as more appropriate to fieldwork than the ParaHIT f.

The inclusion of both a large number of individuals and a large number of malaria-positive patients ensured a high precision in our estimates. Our study was conducted within the routine of a health clinic receiving persons of all ages and allowing the evaluation of the performance of the tests in all persons suspected to have malaria, whatever their history (i.e., recent malaria attack or antimalarial drug intake in the previous days), clinical status (fever or not, severe or mild malaria), demographic characteristics (age and sex), or other factors that may affect the sensitivity and the specificity of a test. Therefore, we believe that the results obtained reflect the performance of the tests in a real situation. Our study showed that a test could be highly valid and reliable but with several inconveniencies for its use in the field, or very practical but with an unacceptably low validity.

One important practical issue is to know how useful an RDT is for diagnosing malaria in the field, taking into account its sensitivity and its specificity. Our study showed that in a setting such as Mbarara out-patient department where about half of the subjects were infected according to the blood smear, positive and negative predictive values were high both for the Paracheck Pf (PPV 91·5%, 95% CI 88·3–93·8; NPV 96·1%, 95% CI 93·2–97·8) and for the ParaHIT f (PPV 91·2%, 95% CI 88·1–93·5; NPV 96·5%, 95% CI 93·4–98·2). In other words, in a setting such as ours, a negative test corresponds in the vast majority of cases to a non-infected individual. The risk of missing an infected individual is, in this case, very small (less than 1 negative test out of 10). The high NPV allows us to confidently diagnose negative-test patients as non-malaria patients. Similar conclusions can be drawn for positive tests, which correspond in the vast majority of cases to an infected individual. A positive test reliably diagnoses a patient as infected by malaria parasites, with a low risk of error (about 1 in 10).

In conclusion, our study showed that the Paracheck Pf dipstick and device were the most appropriate tests for the use in the field, but that ParaHIT f was a good alternative. When laboratory facilities are not available, increased reliance on these rapid tests, which are reliable, cheap, and simple enough to be used by non-laboratory staff, is likely to contribute greatly to an effective control of malaria.

Acknowledgements

We thank the out-patient department staff of Mbarara Teaching Hospital for their invaluable help. Thank you to the local and national health authorities, and to the authorities of Mbarara Teaching Hospital and Mbarara University, who allowed us to conduct this research. Thanks to the laboratories who provided free of charge the test kits necessary to perform this work. This research was financed by Médecins sans Frontières–France.

Table 3. Classification of 5 rapid diagnostic tests for falciparum malaria, according to their general characteristics and the opinion of the readers concerning their ease of use in the field in 742 patients attending Mbarara out-patient department (Mbarara Hospital), south-western Uganda (November 2000–January 2001)

	Score	Paracheck dip.	Paracheck dev.	ParaHIT	BIO P.F.	Malaria Rapid
Characteristics of the test						
Number of invalid tests (%)		1 (0·10%)	2 (0·30%)	4 (0·54%)	0 (0%)	2 (0·30%)
<0·5%	2	2	2		2	2
≥0·5%	1			1		
Number of doubtful tests (%)		0 (0%)	0 (0%)	0 (0%)	3 (0·4%)	19 (2·6%)
<1%	2	2	2	2	2	
>1%	1					1
Shelf-life in months		16	16	11	9	15
>12 months	2	2	2			2
≤12 months	1			1	1	
Storage volume for 1000 tests (in m^3)		0·040	0·075	0·071	0·094	0·059
<0·06	2	2				2
≥0·06	1		1	1	1	
Cost (in US$)		0·55	0·62	0·70	1·30	0·65
Classify from 1 to 5 (5 for the cheapest)	1–5	5	4	2	1	3
Required blood quantity						
5 μL	2	2	2	2		2
25 μL	1				1	
Steps needed						
Direct contact finger/test	2	2		2	2	
Use of a sample applicator pipette	1		1			1
Dropper buffer	2	2	2		2	2
Use of a pipette for the buffer	1			1		
No use of a tube	2		2		2	2
Test to be placed in the tube	1	1		1		
Additional material						
Not required	3		3			3
Required and easy to find	2			2	2	
Required and difficult to find	1	1				
Stability of the result						
>24 h	2	2	2	2		2
≤24 h	1				1	
Opinion of readers						
Quality of the instruction sheet						
Very good	2		2	2	2	2
Good	1	1				
Ease of performance (opinion)						
Very easy	2	2		2		
Easy	1		1		1	1
Ease of box opening						
Easy	3	3	3			3
Difficult	1			1		
Box not provided	2				2	
Ease of bag opening						
Very easy	3				3	
Easy	2	2	2			2
Difficult	1			1		
Score (total = 36)		31	31	23	25	30

dip, dipstick; dev, device.

References

Kilian, A. H., Kabagambe, G., Byamukama, W., Langi, P., Weis, P. & von Sonnenburg, F. (1999). Application of the *Para*Sight-F dipstick test for malaria diagnosis in a district control program. *Acta Tropica*, **72**, 281–293.

Lema, O. E., Carter, J. Y., Nagelkerke, N., Wangai, M. W., Kitenge, P., Gikunda, S. M., Arube, P. A., Munafu, C. G., Materu, S. F., Adhiambo, C. A. & Mukunza, H. K. (1999). Comparison of five methods of malaria detection in the outpatient setting. *American Journal of Tropical Medicine and Hygiene*, **60**, 177–182.

Mount, D. L., Nahlen, B. L., Patchen, L. C. & Churchill, F. C. (1989). Adaptations of the Saker–Solomons test: simple, reliable colorimetric field assays for chloroquine and its metabolites in urine. *Bulletin of the World Health Organization*, **67**, 295–300.

WHO (1996). A rapid dipstick antigen capture assay for the diagnosis of falciparum malaria. WHO Informal Consultation on Recent Advances in Diagnostic Techniques and Vaccines for Malaria. *Bulletin of the World Health Organization*, **74**, 47–54.

Received 12 November 2001; revised 2 January 2002; accepted for publication 9 January 2002