RESEARCH ARTICLE

Bayesian evaluation of the performance of three diagnostic tests for *Plasmodium falciparum* infection in a low-transmission setting in Kilifi County, Kenya [version 1; peer review: 1 approved with reservations]

Marshall M. Mweu¹, Juliana Wambua², Fixtan Njuga², Philip Bejon¹²,³, Daniel Mwanga²

¹School of Public Health, University of Nairobi, Nairobi, Kenya
²KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya
³Centre for Tropical Medicine and Global Health, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK

Abstract

**Background:** Central to the successful elimination of *Plasmodium falciparum* malaria, are tests with superior capability of diagnosing low-density parasitaemias. Empirical evidence on the performance of the commonly available diagnostics (light microscopy (LM), rapid diagnostic tests (RDT) and polymerase chain reaction (PCR)) is needed to better inform case management and surveillance activities within primary health care settings where elimination of *falciparum* malaria is targeted. The objective of this study was to estimate the sensitivity (Se) and specificity (Sp) and predictive values of LM, RDT and PCR tests for *P. falciparum* infection, while evaluating the effect of specific covariates on the accuracy of the tests.

**Methods:** The study enrolled 1,563 children via a cross-sectional survey for asymptomatic malaria and those presenting with symptomatic malaria to the Ngerenya dispensary, Kilifi County between March and December 2014. A Bayesian latent class model (BLCM) was fitted to the participants’ diagnostic data obtained from blood samples that were screened for the presence of *P. falciparum* using the three tests.

**Results:** The PCR assay registered a higher Se (97.6% [92.0; 99.7]) than LM (84.0% [74.8; 91.0]) but similar to RDT (92.2% [84.4; 97.0]). However, the assay showed a similar Sp (98.9% [98.2; 99.4]) to both RDT (99.4% [98.9; 99.7]) and LM (99.5% [99.0; 99.8]). Regarding predictive values, the tests yielded statistically similar estimates of Positive and negative predictive values (PPV and NPV). A serial interpretation of the results of RDT and LM raised the PPVs and NPVs to >98%.

**Conclusions:** LM and RDT tests afford high Se and Sp in a low *P. falciparum* prevalence setting. A serial combination of the tests assures high PPV and NPV estimates. These elements, coupled with the wide
deployment and affordability of the tests, lend the tests useful for guiding clinical care and surveillance activities for *P. falciparum* within elimination settings.

**Keywords**
Bayesian latent class model, *Plasmodium falciparum*, Light microscopy, Rapid diagnostic test, PCR, Test evaluation

This article is included in the KEMRI | Wellcome Trust gateway.

**Corresponding author:** Marshal M. Mweu (marshal@uonbi.ac.ke)

**Author roles:** Mweu MM: Conceptualization, Formal Analysis, Methodology, Writing – Original Draft Preparation; Wambua J: Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing; Njuga F: Methodology, Writing – Review & Editing; Bejon P: Conceptualization, Funding Acquisition, Investigation, Project Administration, Resources, Supervision, Writing – Review & Editing; Mwanga D: Data Curation, Investigation, Project Administration, Resources, Validation

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was supported by the Wellcome Trust [203077]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Copyright:** © 2019 Mweu MM et al. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Mweu MM, Wambua J, Njuga F et al. Bayesian evaluation of the performance of three diagnostic tests for *Plasmodium falciparum* infection in a low-transmission setting in Kilifi County, Kenya [version 1; peer review: 1 approved with reservations] Wellcome Open Research 2019, 4:67 (https://doi.org/10.12688/wellcomeopenres.15204.1)

**First published:** 08 Apr 2019, 4:67 (https://doi.org/10.12688/wellcomeopenres.15204.1)
Introduction

Malaria persists as a leading cause of morbidity and mortality globally\(^1\) and in Kenya, close to 70% of the population is at risk of the disease. *Plasmodium falciparum* is the most preponderant malaria parasite in the country associated with over 99% of malaria infections\(^2\). However, studies report a declining trend of *P. falciparum* prevalence particularly along the Kenyan coastal region\(^3\).

The Kenya national guidelines for diagnosis and treatment of malaria dictate that malaria treatment should be informed by parasitological diagnosis\(^4\). Light microscopic (LM) examination of thin or thick blood smears is held as the standard method for malaria diagnosis\(^5\). The test is inexpensive, generally exhibits high sensitivity (Se) and specificity (Sp) and permits parasite quantification which is a vital attribute in evaluating disease severity and guiding appropriate therapy\(^6\). Nevertheless, the test is fraught with some challenges: it requires good lab equipment and well-trained microscopists, that are often lacking in poor settings, and it may display poor Se especially in cases of low parasitaemia given its detection limit of about 20 parasites per µl of blood\(^7\).

Rapid diagnostic tests (RDTs) are immunochromatographic tests that detect specific parasite antigens\(^5\). The tests target the *P. falciparum*-specific protein i.e. the histidine-rich protein II (HRP–II) or lactate dehydrogenase (LDH)\(^8\). RDTs are hailed for their rapidity, usage simplicity, suitability for use in remote settings with limited equipment and trained staff and display high Se and Sp under field conditions\(^9\). However, their Se may be limited in situations of low parasitaemia\(^10\).

With declining *P. falciparum* transmission rates and thus its prevalence within the country\(^11\), low parasite densities in the population are anticipated that may compromise the Se of LM and RDTs\(^2\). This situation may warrant alternative tests capable of detecting diminished levels of parasitaemia. Molecular-based techniques such as polymerase chain reaction (PCR) assays are touted as being less subjective and affording high Se and Sp in low parasite density settings\(^12\). Nevertheless, as sub-microscopic parasitaemias have unknown clinical importance, the utility of PCRs in guiding national guidelines for clinical case management is vague. Empirical evidence on the performance of these diagnostics in low-transmission settings is thus necessary to better inform management and surveillance efforts for *P. falciparum* malaria.

The diagnostic performance of RDTs has previously been evaluated using LM as a reference test\(^13\). A drawback of this approach is that, given the imperfection of LM, the index tests’ characteristics are subject to bias. Moreover, for index tests (e.g. PCR) presumed to have superior accuracy to the existing reference test, their evaluation based on the reference test is impractical. In the absence of a reliable reference test, latent class models (LCMs) allow for the simultaneous estimation of Se and Sp of two or more tests without any assumption about the underlying true disease status of each individual\(^14\). LCMs can be fit using either maximum likelihood or Bayesian methods\(^15\).

Essentially, Bayesian methods are preferable when observed data are insufficient\(^16\).

In heterogeneous populations, when information on some covariates perceived to influence the performance of index tests is available, stratified estimates of Se and Sp are computable and, reasonably, are more relevant than singular estimates. Of note, estimates of Se and Sp are characteristics specific to a test. However, once the test is applied in any given population, our interest rests on predictive values since the present concern is whether a particular tested individual has/does not have the disease in question given his/her test status. Therefore, the objective of this study was to estimate (within a Bayesian framework) the Se and Sp and predictive values of LM, RDT and PCR diagnostic tests for *P. falciparum* infection, while evaluating the effect of specific covariates on the performance of the tests.

Methods

Study area and population

The study participants comprised asymptomatic children aged <15 years recruited from a cross-sectional survey of the Ngerenyia cohort\(^17\), and patients aged <12 years presenting with malaria symptoms to the Ngerenyia dispensary, Kilifi County between March and December, 2014. These sites are situated within the Kilifi Health and Demographic Surveillance System (KHDSS) area on the Kenyan coast\(^18\). Notably, the KHDSS area spans a population of approximately 260,000 persons in an area of about 891km\(^2\). The area experiences a bimodal rainfall pattern, with long rains coming in the months of April to July and short rains in October and November. Malaria is endemic and transmission occurs throughout the year, with annual entomological inoculation rates ranging between 30 and 100\(^19\).

Sample collection

Blood samples were collected from the study participants for immunological examinations and assessment of asymptomatic and symptomatic *P. falciparum* infections. As part of the cross-sectional survey, the child’s axillary temperature was taken and regardless of their fever status, a finger prick blood sample was taken for rapid malaria diagnosis and a blood film to quantify parasitaemia. At the dispensary, if the child was febrile, a finger prick blood sample was collected for RDT, a slide for LM and an EDTA for PCR. Rapid diagnosis was carried out by trained field workers at the dispensary while slide and EDTA blood samples were stored in cooler boxes and transported to the KEMRI-Wellcome Trust Research laboratory, Kilifi where the LM and PCR tests were run by trained laboratorians. Examiners for parasitaemia using any of the diagnostics were blinded to the results of the other tests. Besides the samples, the participants’ socio-demographic characteristics (area of residence, age, sex and visit date) were captured.

Ethical considerations

Parents/guardians of enrolled children provided written informed consent agreeing to their children’s participation in the study. In addition, assent was secured from the participating minors. Approval for the study was obtained from the Kenya Medical...
Research Institute (KEMRI) Ethical Review Committee (SSC No. 2617).

Target condition
The latent (unobserved) infection status (referred to here as parasitaemia) targeted for detection by the three tests: LM, RDT and PCR, represents a blood sample containing either the live P. falciparum parasite or its HRP-II antigens/products or debris at any concentration level.

Light microscopy (LM)
Malaria microscopy was performed as per standard guidelines. Briefly, thick and thin blood films were stained in 3% Giemsa solution for 45 to 60 minutes and examined under a light microscope at 1000× magnification for malaria parasites. Parasite quantification was achieved by counting the number of malaria parasites per 200 leucocytes. Parasite density per µl of blood was estimated assuming 8000 leucocytes per µl of blood and reported by species i.e. P. falciparum, Plasmodium malariae and Plasmodium ovale. For analytical purposes, P. falciparum densities above zero constituted a positive result; otherwise negative.

Rapid diagnostic test (RDT)
A CareStart™ Malaria RDT kit (Cat No. G0141, AccessBio Inc.) was used to test for the presence of P. falciparum specific HRP-II antigens in the collected blood samples in accordance with instructions contained in the CareStart™ Malaria manual (AccessBio). A total of 5 µl of blood was added into sample wells followed by 60 µl of assay buffer solution added to assay wells. The blood-buffer mixture was then allowed to flow towards the test and control windows. The presence of two colour bands denoted a positive result; one band (the control line) indicating a negative result.

Polymerase chain reaction (PCR) assay
The PCR analysis was conducted as described elsewhere. Briefly, 30 µl of blood was used for DNA extraction using a QIAxtractor machine (QIAGEN, Hilden, Germany). In a subsequent step, the extracted DNA was eluted in a volume of 100 µl, after which 5 µl of DNA was amplified in a thermal cycler (Applied Biosystems ™ 7500 Real-Time PCR System, Applied Biosystems, Foster City, CA). A blood sample of 5 µl was then added to 45 µl of amplification mixture containing TaqMan buffer, 250 µM of each deoxynucleotide triphosphate, 0.125 U AmpliTag Gold polymerase, oligonucleotide primers and TaqMan probe (5’-FAM-AACAATTGGAGGGCAAG-NFQ-MGB-3’) in about 10 minutes at 95°C pre-incubation, PCR amplification was carried out for 45 cycles (15s 95°C, 1 min 60°C) using a TaqMan assay for the highly conserved P. falciparum multicopy 18S ribosomal RNA gene. Quantification was achieved based on the Applied Biosystems 7500 software v2.0.6. The method has a quantification limit of 100 µl, after which 5 µl of DNA was amplified in a thermal cycler (Applied Biosystems ™ 7500 Real-Time PCR System, Applied Biosystems, Foster City, CA). A blood sample of 5 µl was then added to 45 µl of amplification mixture containing TaqMan buffer, 250 µM of each deoxynucleotide triphosphate, 0.125 U AmpliTag Gold polymerase, oligonucleotide primers and TaqMan probe (5’-FAM-AACAATTGGAGGGCAAG-NFQ-MGB-3’). In about 10 minutes at 95°C pre-incubation, PCR amplification was carried out for 45 cycles (15s 95°C, 1 min 60°C) using a TaqMan assay for the highly conserved P. falciparum multicopy 18S ribosomal RNA gene. Quantification was achieved based on the Applied Biosystems 7500 software v2.0.6. The method has a quantification limit of 4.5 parasites/µl of blood. Three negative control wells and seven serial dilutions of DNA extracted from an in vitro culture of the P. falciparum 3D7 parasite strain were included on each plate as standards. Samples with PCR readings beyond zero were considered positive.

Population classification
Organisationally, within the Kenyan health system, a dispensary denotes a primary care unit serving an immediate catchment population. Arguably, therefore, the dispensary and survey data represented a sample drawn from a single target population that formed the basis for derivation of Se and Sp of the three tests.

Statistical analysis
Initially, based on the participants’ visit dates, a dichotomous ‘season’ variable (‘wet’: [April – July, October and November] versus ‘dry’: [January – March, August, September, December]) was generated. A Bayesian latent class model (BLCM) built in OpenBUGS v3.2.25 but called from R software v3.4.3 via the ‘BRugs’ package v0.9-0 was used to infer prevalence, the tests’ characteristics and corresponding predictive values. Notably, the analysis was informed by the guideline for standards for reporting diagnostic accuracy studies that use BLCMs (STARD-BLCM)27. Code used to run this analysis is available as underlying data.

In fitting a BLCM, three assumptions are necessary: (1) the target population should consist of two or more subpopulations with different prevalences, (2) the Se and Sp of the tests under evaluation should be constant across the subpopulations and (3) the tests should be conditionally independent given the disease status. For our situation, the three tests were assumed to be conditionally independent given an individual’s P. falciparum infection status. This is sensible considering that the tests target different aspects of the parasite. As such, statistically, for an individual whose infection status is known, their probability of testing positive (or negative) to one of the tests remains the same regardless of their prior outcomes to the other tests. In order to evaluate the separate effects of ‘season’, ‘sex’ and ‘age’ (dichotomised into <5 yrs and ≥ 5yrs) on the Se and Sp estimates of the tests, we relaxed the assumption on constancy of the tests characteristics by stratifying the single population into subpopulations defined by the levels of the covariates. This allowed for the computation of stratum-specific tests estimates. Hypotheses for the differences between the stratified estimates were evaluated using a Bayesian P-value.
For each covariate, the resulting two subpopulations furnished 14 degrees of freedom sufficient to estimate 14 parameters (stratum-specific Se and Sp of the three tests and two subpopulation prevalences) – suggesting identifiability of the model. Of note, model identifiability can at least be justified if the number of subpopulations ($k$) and tests ($i$) fulfil the equation: $k \geq i(2^{i–1} - 1)$.

Positive and negative predictive values (PPV and NPV respectively) associated with test $i$ and subpopulation $k$ were derived as follows:

$$ ppv = \frac{P_k Se_{ik}}{P_k Se_{ik} + [1 - P_k][1 - Se_{ik}]} $$
$$ npv = \frac{1 - P_k Sp_{ik}}{1 - P_k Sp_{ik} + [1 - P_k][1 - Sp_{ik}]} $$

Non-informative priors ($beta(1,1)$) were used to fit the Bayesian model since no reliable prior information was available for any of the tests parameters. A separate (non-stratified) model ignoring differences in tests Se and Sp across covariate levels was also fitted and the relative goodness of fit for the nested models compared using the Deviance Information Criterion (DIC) (the model with the smaller DIC value being more preferable).

The models were initialised with two Markov Chain Monte Carlo chains with different values. Each chain comprised 70,000 samples, with the first 20,000 being discarded as the burn-in. Convergence of the chains was evaluated by visual appraisal of the time series plots of selected variables and the Gelman-Rubin diagnostic plots. The posterior distribution of the subpopulation prevalences, the Se and Sp of the three tests, as well as the predictive values were reported as the median and the corresponding 95% posterior credible intervals (PCI).

Results

The cross-tabulated counts of the three tests’ outcomes by covariate level are displayed in Table 1. The sample comprised 1563 children, of whom 65.8% ($n = 1029$) were <5 years of age, 47.3% ($n = 739$) were female and 34.5% ($n = 492$) made visits during the dry season.

The stratum-specific estimates of Se and Sp of the three tests for *P. falciparum* are presented in Table 2. There were no detectable significant differences between the covariate-stratified tests’ estimates as indicated by the Bayesian $P$-value. Furthermore, the non-stratified model gave better fit ($DIC = 41.7$) to the data than any of the covariate-stratified models ($DICs = [72.9; 72.0; 69.9]$) and was thus utilised for subsequent analyses.

Results of the estimates of Se and Sp of the three tests together with their respective predictive values and prevalence of *P. falciparum* are shown in Table 3. The PCR assay recorded a higher Se (97.6; 95% PCI [92.0; 99.7]) than LM (84.0; 95% PCI [74.8; 91.0]) but similar to RDT (92.2; 95% PCI [84.4; 97.0]). Nonetheless, the assay registered a similar Sp (98.9; 95% PCI [98.4; 99.9]) to both RDT (99.4; 95% PCI [98.9; 99.7]) and LM (99.5; 95% PCI [99.0; 99.8]). As for predictive values, the tests had statistically similar estimates of PPV and NPV. Since PPVs were comparably lower than NPVs, in a bid to bolster their estimates, a serial interpretation of the results of RDT and LM led to a considerable improvement in PPV (99.9; 95% PCI [99.8;100.0]) at a negligible expense of NPV (98.7 95% PCI [97.9; 99.2]).

Discussion

Using latent class analysis we have estimated the accuracy of LM, RDT and PCR tests for the diagnosis of *P. falciparum* infection along with their associated predictive values. Enoe et al. contend...

---

### Table 1. Cross-classified results by stratum for rapid diagnostic test (RDT), light microscopy (LM) and polymerase chain reaction (PCR) tests for diagnosis of *P. falciparum* infection in the study population in Ngerenya, Kilifi County, Kenya.

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Tests outcomes combinations (RDT; LM; PCR)</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td>Single population</td>
<td>67</td>
<td>1</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;5yrs</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>≥5yrs</td>
<td>43</td>
<td>0</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>Male</td>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td>Season</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dry</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>Wet</td>
<td>38</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 2. Stratum-specific estimates of sensitivity and specificity of rapid diagnostic test (RDT), light microscopy (LM) and polymerase chain reaction (PCR) tests for *P. falciparum* infection and a Bayesian *P*-value for the difference between the stratified estimates.

<table>
<thead>
<tr>
<th>Test parameter</th>
<th>Covariate</th>
<th>Bayesian <em>P</em>-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;5yrs</td>
<td>≥5yrs</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;RD&lt;/sub&gt;</td>
<td>Estimate (95% PCI)&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Estimate (95% PCI)</td>
</tr>
<tr>
<td></td>
<td>89.8 (74.4; 97.7)</td>
<td>92.0 (81.9; 97.7)</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;LM&lt;/sub&gt;</td>
<td>77.6 (60.5; 90.0)</td>
<td>86.3 (75.0; 94.1)</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;PCR&lt;/sub&gt;</td>
<td>93.3 (79.5; 99.1)</td>
<td>98.4 (91.6; 99.9)</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;RD&lt;/sub&gt;</td>
<td>99.5 (98.9; 99.9)</td>
<td>99.1 (97.9; 99.7)</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;LM&lt;/sub&gt;</td>
<td>99.6 (99.0; 99.9)</td>
<td>99.3 (98.2; 99.8)</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;PCR&lt;/sub&gt;</td>
<td>98.9 (98.1; 99.5)</td>
<td>98.7 (97.4; 99.6)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td>Estimate (95% PCI)</td>
<td>Estimate (95% PCI)</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;RD&lt;/sub&gt;</td>
<td>94.1 (81.2; 99.3)</td>
<td>89.4 (78.2; 96.2)</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;LM&lt;/sub&gt;</td>
<td>84.8 (69.5; 94.5)</td>
<td>81.9 (69.4; 91.0)</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;PCR&lt;/sub&gt;</td>
<td>97.4 (87.2; 99.9)</td>
<td>96.0 (87.0; 99.5)</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;RD&lt;/sub&gt;</td>
<td>99.6 (99.0; 99.9)</td>
<td>99.0 (98.2; 99.6)</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;LM&lt;/sub&gt;</td>
<td>99.5 (98.8; 99.9)</td>
<td>99.4 (98.7; 99.8)</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;PCR&lt;/sub&gt;</td>
<td>98.7 (97.6; 99.4)</td>
<td>99.0 (98.0; 99.6)</td>
</tr>
<tr>
<td><strong>Season</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dry</td>
<td>Wet</td>
</tr>
<tr>
<td></td>
<td>Estimate (95% PCI)</td>
<td>Estimate (95% PCI)</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;RD&lt;/sub&gt;</td>
<td>97.6 (87.8; 99.9)</td>
<td>86.8 (74.5; 94.8)</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;LM&lt;/sub&gt;</td>
<td>80.7 (65.6; 91.5)</td>
<td>84.8 (72.3; 93.4)</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;PCR&lt;/sub&gt;</td>
<td>97.6 (88.0; 99.9)</td>
<td>95.8 (86.3; 99.5)</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;RD&lt;/sub&gt;</td>
<td>99.2 (98.1; 99.8)</td>
<td>99.4 (98.7; 99.8)</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;LM&lt;/sub&gt;</td>
<td>99.6 (98.8; 100.0)</td>
<td>99.4 (98.7; 99.8)</td>
</tr>
<tr>
<td><em>S</em>&lt;sub&gt;PCR&lt;/sub&gt;</td>
<td>99.0 (97.8; 99.7)</td>
<td>99.8 (97.9; 99.4)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Median estimates  
<sup>b</sup>Posterior credible interval  
<sup>c</sup>Value is considered significant if it lies outside the range 0.025; 0.975

that a BLCM framework permits derivation of true estimates of index tests devoid of classification errors that may be introduced by the utilisation of an imperfect reference test. Thus, the findings of this study can be considered readily generalisable to other settings with similar *P. falciparum* infection burden. Since malaria transmission dynamics have been shown to differ by age<sup>3</sup>, season<sup>31</sup> and sex<sup>32</sup>, it is conceivable that the accuracy of malaria diagnostics may be influenced by these covariates. To test for this, the Se and Sp of the three tests were allowed to vary by the aforementioned covariate levels. However, it was shown that neither of the tests’ characteristics differed significantly across the levels of any of the covariates. This implies that the performance of the tests is not influenced by either the prevailing season or the age and sex of the presenting patient. Nonetheless, in São Tomé and Príncipe, LM was reported to have a lower Se in afebrile under-five children suggesting the inadequacy of the test in detection of low-density parasitaemias<sup>33</sup>.

In the present study setting, the prevalence of *P. falciparum* was estimated to be 5.6% [95% PCI 4.5; 6.8], suggesting a low transmissibility of the parasite in the population. Accordingly,
The strong confidence realised in a negative test outcome is especially fundamental in a low-transmission setting where the preponderance of low-density infected individuals calls for their accurate detection not only to inform treatment but also to guide successful *P. falciparum* malaria elimination efforts. Importantly, low-density parasitaemic individuals present as potential reservoirs of infection to uninfected mosquitoes so that, if undetected, transmission may be sustained silently.84.

**Conclusions**

Using a Bayesian approach, we have derived the Se and Sp of LM, RDT and PCR for the diagnosis of *P. falciparum* infection as well as their associated predictive values. It was shown that the PCR assay’s Se was significantly higher than that of LM but similar to RDT. Nevertheless, the Sp estimates of the three tests were similar. Furthermore, the three tests produced comparable estimates of predictive values. In an elimination setting, a serial interpretation of the results of both RDT and LM raised the confidence of PPV and NPV. In particular, these estimates signify a reduced confidence in a positive compared to a negative test result ascribable to the low *P. falciparum* prevalence. A serial interpretation of the results of both RDT and LM raised the confidence to >98% in both positive and negative test outcomes. Consequently, in this low-transmission setting where false positives are increasingly expected, the most optimal testing strategy should be one that has all individuals initially screened by the more sensitive RDT test, with any resulting positives followed up with the more specific LM. Only those individuals positive to both tests should be eligible for treatment. This multiple-test approach is pivotal to reducing the risk of parasite resistance that can occur when individuals are unnecessarily subjected to artemisinin therapy83. The rationale for employing the RDT-LM test combination owes to the wide deployment and affordability of the tests in most primary care settings83 granting them promise towards informing clinical care and surveillance activities aimed at eliminating *falciparum* malaria. Moreover, as quantification of parasitaemias is central to the management of severe malaria and assessment of treatment response84, in this respect, LM affords an added merit. By contrast, as PCRs demand hefty investment in equipment and reagents as well as highly trained personnel, their potential for routine use in low-resource field settings is limited.

The raw dataset for the study is stored under restricted access since it contains sensitive participant information. Notwithstanding, accessibility is possible upon placing a formal request to our Data Governance Committee (dgc@kemri-wellcome.org). The replication data and analysis scripts for this manuscript are available from the Harvard Dataverse.
Harvard Dataverse: Replication data for: Bayesian evaluation of the performance of three diagnostic tests for Plasmodium falciparum infection in a low-transmission setting in Kilifi County, Kenya. https://doi.org/10.7910/DVN/Z5RBBT

This project contains the following underlying data:
- BLCM_malaria_tests_code.R (Bayesian Latent Class R code for malaria tests evaluation)
- Ngerenya_tests_data.tab (Study dataset)

Data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

References


Open Peer Review

Gillian Stresman

Department of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK

Overall comments:

This manuscript describes use of a latent class analysis to estimate the performance of three different diagnostic tools for malaria. The models and statistical methods are well described and the results are an interesting addition to the thinking about malaria diagnostic performance in a low transmission setting. However, I can only approve with reservations. As this paper is currently written, important nuances with the interaction between diagnostic test sensitivity and parasite densities (e.g. clinical infections vs infections identified in the community), sufficient details on the study population for the results to be interpreted accordingly, and how this relates to the broader generalizations that can be made should be addressed.

Specific comments:

- The background argument to justify the study should be reframed slightly to better reflect the nuances with malaria diagnostics. The diagnostic performance of RDTs and LM has been assessed vs PCR in many research settings and is now considered as the gold standard, including some of the references cited by the authors. The text suggests that this is not the case with LM being the main comparator.

- Similarly, the authors state that PCR has presumed superior accuracy, but PCR is known to have superior sensitivity compared to RDTs and LM. The key question with PCR, as the authors state, being does that matter in a clinical/elimination setting? There is evidence that standard PCR methods are not perfect with the harder we look, the more parasites we’ll find. This argument should be included here as a better justification of the approach taken.

- Samples are based on suspected clinical cases under 12 years of age and children 15 years of age sampled in the community. These populations are inherently different in terms of their likely parasite densities in infected individuals (with high parasitemia related to the probability of clinical symptoms) and demographically in terms of factors associated with care-seeking behavior. This is
a critical strata that is currently missing to assess any effects on the Sp/Se of the tests and may contribute to the lack of effects observed.

- As a related point, in the discussion, the authors state that Se/Sp of RDT and PCR have been found to be similar in the diagnosis of severe malaria in children in a clinical setting. This is not surprising as clinical cases of malaria are likely to have parasite densities well above the limit of detection of RDTs. This is very different than detection infections with expected low parasite densities. As above, the critical implications of the populations being studied on the interpretation of the results in this context should be more carefully expressed and acknowledged.

- Parasite densities in children are expected to be higher and more likely to be detectable using routine diagnostic tools (see Wu et al., 2015\(^1\)). It is therefore critical to interpret the resulting Se/Sp estimates as detection of malaria infections in children. Results from this study are not generalizable to adult populations.

- In the introduction the authors rightly claim that light microscopy is the standard for clinical malaria diagnosis and that its quality is highly variable. It might clarify these important nuances by distinguishing between expert microscopy, which has high Se and Sp, and routine microscopy, which is known to have a detection limit 10-fold higher than that stated and more similar to that of RDTs.

- I'm not aware of any RDTs for \(Pf\) that detect LDH only. Do the authors mean detect HRP-II and LDH?

- In discussing RDTs it would be helpful to follow the same format as per microscopy and include the limit of detection for parasite densities for the routinely used tests, which is available via the WHO website.

- In terms of the definition of the ‘season’ variable, I’m not entirely clear about which population this is referring to. Is this for the clinical cases only? What time period of suspected cases were eligible for inclusion? In the methods section this is not clear. Similarly, for the community samples, the authors state that it’s a cross-sectional survey, which implies a single time point but also mention cohort, which implies multiple timepoints of sampling per individual. Was the community work part of a panel study where repeated cross-sectional surveys were conducted in the same population? Please clarify in both the methods and the definition of seasonality.

- “In the present study setting, the prevalence of \(P. falciparum\) was estimated to be 5.6%…” - what malaria metric was used to derive this estimate?

- The authors are correct in that HRP-II gene deletion in parasites will impact the Se, but so will circulating antigen levels that are below the limit of detection. It may also take time for antigen to build up to detectable levels, etc. There are lot of reasons why RDT Se is imperfect. A more thorough discussion would be helpful to highlight these important nuances.

- False positive test results are equally of concern in high-prevalence settings (and arguably more likely given residual HRP-II).
● LDH tests are also much less sensitive in detecting infections so may not be the best choice for monitoring treatment efficacy.

● False positive PCR results may also be linked to contamination in the lab process.

● Although routine use of PCR in most malaria endemic settings is impractical as the authors suggest, in low transmission/pre-elimination settings would there be a role for reference labs to be set up for confirmatory testing?

● In the introduction the authors state that low density/PCR detectable infections are of unconfirmed importance, but the conclusion is that they are essential to remove as they are a potential source of onward transmission? This seems contradictory. In the introduction are the authors referring to the current WHO position? And in the conclusion the position from some of the research community?

References

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: malaria epidemiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.