

Maria Regina Fernandes de Oliveira^I

Almério de Castro Gomes^{II}

Cristiana Maria Toscano^{III}

QUADAS and STARD: evaluating the quality of diagnostic accuracy studies

ABSTRACT

OBJECTIVE: To compare the performance of two approaches, one based on the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) and another on the Standards for Reporting Studies of Diagnostic Accuracy (STARD), in evaluating the quality of studies validating the OptiMal® rapid malaria diagnostic test.

METHODS: Articles validating the rapid test published until 2007 were searched in the Medline/PubMed database. This search retrieved 13 articles. A combination of 12 QUADAS criteria and three STARD criteria were compared with the 12 QUADAS criteria alone. Articles that fulfilled at least 50% of QUADAS criteria were considered as regular to good quality.

RESULTS: Of the 13 articles retrieved, 12 fulfilled at least 50% of QUADAS criteria, and only two fulfilled the STARD/QUADAS criteria combined. Considering the two criteria combination (≥ 6 QUADAS and ≥ 3 STARD), two studies (15.4%) showed good methodological quality. The articles selection using the proposed combination resulted in two to eight articles, depending on the number of items assumed as cutoff point.

CONCLUSIONS: The STARD/QUADAS combination has the potential to provide greater rigor when evaluating the quality of studies validating malaria diagnostic tests, given that it incorporates relevant information not contemplated in the QUADAS criteria alone.

DESCRIPTORS: Evaluation Studies as Topic. Diagnosis. Validity of Tests. Reproducibility of Results. Review Literature as Topic.

^I Área de Medicina Social. Faculdade de Medicina. Universidade de Brasília. Brasília, DF, Brasil

^{II} Faculdade de Saúde Pública. Universidade de São Paulo. São Paulo, SP, Brasil

^{III} Instituto de Patologia Tropical e Saúde Pública. Universidade Federal de Goiás. Goiânia, GO, Brasil

Correspondence:

Maria Regina Fernandes de Oliveira
Núcleo de Medicina Tropical
Universidade de Brasília
Campus Universitário Darcy Ribeiro
Caixa Postal 04517 – Asa Norte
70904-900 Brasília, DF, Brasil
E-mail: regifer@unb.br

Received: 4/6/2010
Approved: 8/25/2010

Article available from: www.scielo.br/rsp

INTRODUCTION

New technologies, especially those related to disease diagnosis, must be validated by means of accuracy evaluations. This involves comparing the new test to other, established ones, which are regarded as a gold-standard. Such evaluation is essential to guide the use of a given diagnostic test, especially in the context of widespread use by public health services. The quality and methodological rigor of evaluation studies, as well as the quality of the data obtained, depend on factors that must also be measured and considered.

The technique traditionally used for malaria diagnosis is microscopy. Though inexpensive, this technique requires the presence of trained and experienced professionals. Beginning in the 1990's, rapid tests (RT) were introduced as an alternative to microscopy for malaria diagnosis. Different diagnostic tests are currently on the market.²¹ RTs rely on immunochromatographic methods, and can be administered in about 15 minutes by persons with minimal technical training and using kits that do not require electricity or special equipment.^{12,21}

TR are an effective alternative for malaria diagnosis, for in addition to being easy to implement, their accuracy can be similar to that of microscopy in a number of settings.²¹ The high initial cost of RT is one of the major impediments to its widespread adoption.²¹

OptiMal® is one of the RT registered and validated in Brazil. This test was purchased by the Brazilian Ministry of Health in 2006 for use within the Sistema Único de Saúde (SUS – Unified Health Care System).^{1,a} There are countless validation studies of OptiMal® published in the literature. Different studies deal with populations of endemic and susceptible areas, travelers, symptomatic and asymptomatic populations, and with different clinical aspects of *Plasmodium falciparum* malaria. Determining the quality of these studies using standardized methodology will be fundamental to inform any decisions regarding their use in Brazil.

Two instruments are widely in use in the scientific literature for evaluating the quality of studies validating diagnostic tests: the Standards for Reporting Studies of Diagnostic Accuracy (STARD),³ comprising 25 criteria, and the Quality Assessment of Diagnostic Accuracy Studies (QUADAS),¹⁹ comprising 14 criteria. A number of criteria are common to the two studies.

STARD is an instrument aimed at researchers and editors. It was devised by a group of editors with the purpose of evaluating the quality of articles by simple checking of each of the items in the score, and of guiding authors when elaborating scientific reports.³ QUADAS is intended as an instrument for assessing the quality of previously published studies, especially in the context of systematic literature reviews. This instrument was commissioned by the United Kingdom's NHS R&D Health Technology Assessment Programme (HTA).¹⁹

QUADAS and STARD were created with different aims and applications. Researchers have discussed the need to introduce modifications or combinations of parameters to potentiate the use of these instruments as well as to improve the evaluation of validation studies.^{2,20} Although the aim of STARD is not to evaluate studies included in systematic reviews, the introduction of three of its criteria into such evaluations has been suggested; these are three special items that can provide essential information when evaluating epidemiological studies and methods, and which are absent from QUADAS. We believe that QUADAS, an instrument that has been validated, is considered easy to use,²⁰ and is widely employed in systematic reviews of validation studies, could be improved by the addition of items pertaining to sampling, estimate precision, and the characteristics of study populations. This is important given that validation studies must be representative, precise, and have good external validity to the population of interest.

These issues were discussed during the elaboration of the instrument, but the corresponding criteria were not included in the final document.¹⁹

The present study aimed to compare two different approaches based on the QUADAS and STARD criteria in their ability to assess the quality of validation studies of the rapid malaria test, irrespective of estimates of test accuracy reported by each of the studies evaluated.

METHODS

Validation studies of the OptiMal® RT were obtained from the scientific literature. The bibliographic survey was carried out in December 2007.

We surveyed the available literature using the following inclusion criteria: 1) Studies must use microscopy as a gold-standard; and 2) The study population must comprise patients with clinical signs of malaria, symptomatic, and living in endemic areas, regardless of age group. The first inclusion criterion was used to decide whether the article would be read in full, and the second, to decide whether it would be definitively included in the study. Studies evaluating the accuracy of OptiMal® were selected from within the Medline database using the PubMed search engine. The following key words were used in the search: “evaluation” and “malaria” and “rapid tests” and “diagnosis” (first search) and “OptiMal®” and “malaria” and “diagnosis” (second search). Secondary searches were also carried out in the SciELO and Lilacs databases using the same terms. There were no limitations as to year of publication.

We excluded studies carried out exclusively with patients in specific population subgroups, such as pregnant women, children, or severe malaria patients.

The selected articles were read and analyzed according to a combination of 12 criteria from QUADAS and three from STARD.

QUADAS comprises 14 criteria, 12 of which were considered. The considered criteria are as follows: 1) Was the spectrum of patients representative of the patients who will receive the test in practice?; 2) Were selection criteria clearly described?; 3) Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?; 4) Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?; 5) Did patients receive the same reference standard regardless of the index test result?; 6) Was the execution of the index test described in sufficient detail to permit replication of the test?; 7) Was the execution of the reference standard described in sufficient detail to permit its replication?;

^a Ministério da Saúde. Secretaria de Vigilância em Saúde. Guia de Vigilância Epidemiológica. Série A. Normas e Manuais Técnicos. Brasília: Ministério da Saúde, 2006.

8) Were the index test results interpreted without knowledge of the results of the reference standard?; 9) Were the reference standard results interpreted without knowledge of the results of the index test?; 10) Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?; 11) Were uninterpretable/intermediate test results reported? 12) Were withdrawals from the study explained? Each item must be answered as “yes,” “no,” or “unclear”; the latter should be used in case the available information is deemed insufficient to make a yes/no call. The instrument can be used in its entirety or not; the researcher should select the items considered to be relevant to the index test.¹⁹

We did not consider the criterion “does the gold-standard correctly classify the disease?,” since one of the criteria for inclusion of articles into our study was use of the thick smear as a gold-standard, and thus maintaining this item in the QUADAS scale would be redundant. For the same reason, we did not consider the criterion “is the gold-standard independent of the index test?” since we knew beforehand that the tests are distinct technologies, and there was therefore no reason to categorize this item. The QUADAS instrument does not determine *a priori* the scores for defining quality; it is up to the researcher to decide which cutoff point to use. We therefore considered the fulfilling of six to eight criteria (“yes” answers) as the median cutoff point for defining regular and good studies, and the 75% cutoff point – at least nine criteria – as the definition of a good quality study.

Of the 25 STARD criteria, three were selected as being absent from QUADAS and pertaining to the representativeness and precision of the study sample, both of which are fundamental to the evaluation of quality of epidemiological studies. The remaining STARD items are already included, directly or indirectly, in QUADAS. The three criteria considered were: Item 5 – The sampling process is described; Item 21 – Sensitivity and specificity results are reported with their respective confidence intervals (CI); and Item 15 – Clinical and demographic characteristics of patients are reported. The answers to these three items were dichotomous (yes/no). Good-quality studies should fulfill all three STARD criteria.

Since some of the QUADAS criteria could be interpreted differently by different researchers, we defined parameters to be considered when evaluating the three following criteria: 1) Were selection criteria (of cases) clearly described? – The sample was considered as well-defined in the methodology section when reporting the criteria used for the inclusion of cases (for example: patient with suspected malaria, presenting with acute febrile syndrome) and informing the provenance and recruitment of cases. 2) Was the execution of the index test described in enough detail to allow for its replication? – we considered a description as adequate when

including the techniques used for administering and reading the RT. 3) Was the execution of the index test described in enough detail to allow for its replication? – We considered a description appropriate when the article described the techniques used for coloring and reading the thick smear test.

RESULTS

Our literature search retrieved a total of 254 references, 11 of which were duplicates. All abstracts were read, and 30 articles were selected for full examination, all of which validated the OptiMal® test using microscopy as a gold-standard (first inclusion criterion). Of these, 29 were read in full; we were unable to obtain the full article for one of the 30 abstracts.

Thirteen studies fulfilled all requirements of the second inclusion criterion (Table 1).^{1,4,5,7-10,13-18} Tables 2 and 3 present the results of the evaluation of selected articles according to the selected QUADAS and STARD criteria. Four QUADAS criteria were fulfilled by all studies evaluated: 1) representative spectrum of patients; 2) clear description of selection criteria; 3) entire sample or subsample diagnosed by the gold-standard; and 4) patients received the same test as a gold-standard, regardless of the result of the index test. A smaller number of articles fulfilled the selected STARD criteria, the criterion regarding the confidence intervals being the one most frequently fulfilled (seven of 13 articles; Table 2).

Table 1. Selected OptiMal® validation studies, according to first author, year of publication, study site, and number of patients included. 2007.

	First Author	Year of publication	Study site	Number of patients included
1	Palmer CJ	1998	Honduras	202
2	Cooke AH	1999	Gambia	409
3	Aslan G	2001	Turkey	190
4	Ferro BE	2002	Colombia	112
5	Londoño B	2002	Colombia	107
6	Iqbal J	2003	Pakistan	930
7	Pattanasin S	2003	Thailand	271
8	Singh N	2003	India	155
9	Kolaczinski J	2004	Afghanistan	499
10	Soto-Tarazona A	2004	Peru	72
11	Gonzalez-Ceron L	2005	Mexico	893
12	Van Den Broek I	2006	Colombia	896
13	Ratsimbaoa A	2007	Madagascar	194

Table 2. Combination of QUADAS and STARD criteria and classification of accuracy studies of the OptiMal® RT selected from the literature, according to number of studies in each response category. 2007.

Criteria	Number of articles in each category		
	Yes	No	Unclear
QUADAS			
Was the spectrum of patients representative of the patients who will receive the test in practice?	13	-	-
Were selection criteria clearly described?	13	-	-
Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	7	-	6
Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	13	-	-
Did patients receive the same reference standard regardless of the index test result?	13	-	-
Was the execution of the index test described in sufficient detail to permit replication of the test?;	6	7	-
Was the execution of the reference standard described in sufficient detail to permit its replication?	2	11	-
Were the index test results interpreted without knowledge of the results of the reference standard?	8	-	5
Were the reference standard results interpreted without knowledge of the results of the index test?	7	-	6
Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	-	-	13
Were uninterpretable/intermediate test results reported?	4	9	-
Were withdrawals from the study explained? ^a	2	-	-
STARD			
The sampling process is described	2	11	-
Sensitivity and specificity results are presented with their respective confidence intervals	7	6	-
The demographic characteristics of patients are described	5	8	-

^a there were no losses in 11 studies, therefore this criterion was not applicable

QUADAS: Quality Assessment of Diagnostic Accuracy Studies

STARD: Standards for Reporting Studies of Diagnostic Accuracy

None of the articles fulfilled nine of the 12 QUADAS criteria, and 12 of the 13 articles were categorized as positive in at least 50% of criteria. Five studies failed to fulfill all three STARD criteria (Table 3).

Using a cutoff of six positive responses to the 12 QUADAS criteria and all three QUADAS criteria, two studies (15.4%) were considered as of good methodological quality, regardless of the estimated accuracy reported. Two studies carried out in Colombia fulfilled eight QUADAS criteria (67% fulfillment) and the three STARD criteria.

The number of selected articles using the proposed combination of criteria ranged from two to eight, depending on the number of STARD criteria required in the cutoff point, even when maintaining a median cutoff of 50% of QUADAS criteria (Table 3).

DISCUSSION

The two instruments – QUADAS and STARD – represent advancement in scientific knowledge in that they allow for a systematic evaluation of published validation studies.

QUADAS is a flexible instrument that allows for the exclusion of any of its criteria.^{19,20}

The criterion “were all losses from the study explained?” did not add discriminatory capacity to the evaluation: losses were observed in only two studies, and were all explained. The “does not apply” category does not exist in QUADAS, and should be added specifically for this item. Similar problems were encountered for the criterion “were uninterpretable/intermediate test results reported?” which would be useful in cases of results expressed as a continuous scale or which included the possibility of classifying results as uninterpretable. It is likely that many of the studies for which this criterion was classified as a “no” are cases in which it is not applicable. Similar considerations regarding the absence of adequate categorization of these two items from the form were reported in a study that evaluated and validated QUADAS.²⁰ These two criteria also showed the lowest agreement in the QUADAS validation study,²⁰ as well as in a review of psychometric instruments,¹¹ perhaps reflecting difficulties with the administration of the questionnaire.

Parameters should be established for the evaluation of criteria judging the selection of study subjects for

Table 3. Selected OptiMal® validation studies, according to first author, year of publication, study site, and number of “yes” answers to 12 QUADAS and three STARD criteria. 2007.

First Author	Year of publication/release	Study site	QUADAS – “yes”	STARD – “yes”
Londoño B	2002	Colombia	8	3
Van Den Broek I	2006	Colombia	8	3
Pattanasin S	2003	Thailand	7	2
Ratsimbaoa A	2007	Madagascar	7	2
Gonzalez-Ceron L	2005	Mexico	7	1
Palmer CJ	1998	Honduras	7	1
Ferro BE	2002	Colombia	6	1
Singh N	2003	India	6	1
Iqbal J	2003	Pakistan	8	0
Soto-Tarazona A	2004	Peru	7	0
Cooke AH	1999	Gambia	7	0
Aslan G	2001	Turkey	6	0
Kolaczinski J	2004	Afghanistan	4	0

QUADAS: Quality Assessment of Diagnostic Accuracy Studies

STARD: Standards for Reporting Studies of Diagnostic Accuracy

validation studies and the description of both index and gold-standard tests. The researcher must define *a priori* which information will be sufficient to obtain a “yes” response in these items. Likewise, the criterion “detailed description of diagnostic tests” can mean different things to different evaluators, and again *a priori* standardization will be necessary, especially in the case of multiple reviewers.

We expect that articles following STARD criteria will be better classified according the QUADAS instrument, given that the former provides guidelines and information that are useful for publication of validation findings. Adding items from outside QUADAS and that complement this instrument by responding to specific questions is a strategy that is recommended in the QUADAS validation study itself.²⁰ Clearer knowledge of what is to be evaluated and of the purpose of the information obtained, will lead to a better evaluation of the studies under review.

A systematic review by Fontela et al,⁶ focusing on the diagnosis of malaria, tuberculosis and HIV, highlighted the complementarity of the two instruments in

determining the quality of published articles. Whereas STARD allows one to check the information that ideally should be contained in published validation articles, QUADAS allows one to evaluate the quality of the published information.

The use of instruments to assess the quality of published articles is an increasingly encouraged and useful practice for evidence analysis, especially in the context of systematic reviews and meta-analyses. The use of such instruments, however, does not substitute for a careful and judicious qualitative analysis of the concepts and methods in the study. This is a key task of the researcher when carrying out a literature review.

In conclusion, the QUADAS and STARD instruments are important means to support and substantiate clinical and public health decision-making regarding the use of diagnostic tests. Its combined use has the potential to confer greater rigor to the evaluation of quality of published articles validating malaria diagnostic tests, due to its incorporation of relevant information not contemplated by the use of QUADAS alone. The flexibility of both instruments allows them to be adapted to the purpose of each study.

REFERENCES

- Aslan G, Ulukanligil M, Seyrek A, Erel O. Diagnostic performance characteristics of rapid dipstick test for *Plasmodium vivax* malaria. *Mem Inst Oswaldo Cruz*. 2001;96(5):683-6. DOI:10.1590/S0074-02762001000500018
- Bachmann LM, ter Riet G, Weber WE, Kessels AG. Multivariable adjustments counteract spectrum and test review bias in accuracy studies. *J Clin Epidemiol*. 2009;62(4):357-61. DOI:10.1016/j.jclinepi.2008.02.007
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem*. 2003;49(1):7-18. DOI: 10.1373/49.1.7

4. Cooke AH, Chiodini PL, Doherty T, Moody AH, Ries J, Pinder M. Comparison of a parasite lactate dehydrogenase-based immunochromatographic antigen detection assay (OptiMAL) with microscopy for the detection of malaria parasites in human blood samples. *Am J Trop Med Hyg.* 1999;60(2):173-6.
5. Ferro BE, Gonzalez IJ, Carvajal F, Palma GI, Saraiva NG. Performance of OptiMAL(R) in the diagnosis of *Plasmodium vivax* and *Plasmodium falciparum* infections in a malaria referral center in Colombia. *Mem Inst Oswaldo Cruz.* 2002;97(5):731-5. DOI:10.1590/S0074-02762002000500025
6. Fontela PS, Pant Pai N, Schiller I, Dendukuri N, Ramsay A, Pai M. Quality and reporting of diagnostic accuracy studies in TB, HIV and malaria: evaluation using QUADAS and STARD standards. *PLoS One.* 2009;4(11):e7753. DOI:10.1371/journal.pone.0007753
7. Gonzalez-Ceron L, Rodriguez MH, Betanzos AF, Abadía A. Eficacia de una prueba rápida para el diagnóstico de *Plasmodium vivax* en pacientes sintomáticos de Chiapas, México. *Salud Publica Mex.* 2005;47(4):282-7. DOI:10.1590/S0036-36342005000400005
8. Iqbal J, Muneer A, Khalid N, Ahmed MA. Performance of the OptiMAL test for malaria diagnosis among suspected malaria patients at the rural health centers. *Am J Trop Med Hyg.* 2003;68(5):624-8.
9. Kolaczinski J, Mohammed N, Ali I, Ali M, Khan N, Ezard N, et al. Comparison of the OptiMAL rapid antigen test with field microscopy for the detection of *Plasmodium vivax* and *P. falciparum*: considerations for the application of the rapid test in Afghanistan. *Ann.Trop.Med.Parasitol.* 2004;98(1):15-20. DOI:10.1179/000349804225003127
10. Londoño B, Carmona J, Blair S. Comparación de los métodos OptiMAL y gota gruesa para el diagnóstico de malaria en una zona endémica sin epidemia. *Biomedica.* 2002;22(4):466-75.
11. Mann R, Hewitt CE, Gilbody SM. Assessing the quality of diagnostic studies using psychometric instruments: applying QUADAS. *Soc Psychiatry Psychiatr Epidemiol.* 2009;44(44):300-7. DOI:10.1007/s00127-008-0440-z
12. Moody A. Rapid diagnostic tests for malaria parasites. *Clin Microbiol Rev.* 2002;15(1):66-78. DOI:10.1128/CMR.15.1.66-78.2002
13. Palmer CJ, Lindo JF, Klaskala WI, Quesada JA, Kaminsky R, Baum MK, et al. Evaluation of the OptiMAL test for rapid diagnosis of *Plasmodium vivax* and *Plasmodium falciparum* malaria. *J Clin Microbiol.* 1998;36(1):203-6.
14. Pattanasin S, Proux S, Chompasuk D, Luwiradaj K, Jacquier P, Looareesuwan S, et al. Evaluation of a new Plasmodium lactate dehydrogenase assay (OptiMAL-IT) for the detection of malaria. *Trans R Soc Trop Med Hyg.* 2003;97(6):672-4.
15. Ratsimbaoa A, Randriamanantena A, Raheerinjafy R, Rasoarilalao N, Ménard D. Which malaria rapid test for Madagascar? Field and laboratory evaluation of three tests and expert microscopy of samples from suspected malaria patients in Madagascar. *Am J Trop Med Hyg.* 2007;76(3):481-5.
16. Singh N, Valecha N, Nagpal AC, Mishra SS, Varma HS, Subbarao SK. The hospital- and field-based performances of the OptiMAL test, for malaria diagnosis and treatment monitoring in central India. *Ann Trop Med Parasitol.* 2003;97(1):5-13. DOI:10.1179/000349803125002544
17. Soto Tarazona A, Solari Zerpa L, Mendonza Requena D, Llanos-Cuentas A, Magill A. Evaluation of the rapid diagnostic test OptiMAL for diagnosis of malaria due to *Plasmodium vivax*. *Braz J Infect Dis.* 2004;8(2):151-5. DOI:10.1590/S1413-86702004000200005
18. Van den Broek I, Hill O, Gordillo F, Angarita B, Hamade P, Counihan H, et al. Evaluation of three rapid tests for diagnosis of *P. falciparum* and *P. vivax* malaria in Colombia. *Am J Trop Med Hyg.* 2006;75(6):1209-15.
19. Whiting P, Rutjes AW, Dinnes J, Reitsma J, Bossuyt PM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess.* 2004;8(25):iii,1-234.
20. Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol.* 2006;6:9. DOI:10.1186/1471-2288-6-9
21. World Health Organization. Malaria diagnosis: new perspectives. Geneva: WHO Graphics; 2000.

Article based on the Doctoral thesis of Oliveira MRF, presented to the Faculdade de Saúde Pública da Universidade de São Paulo in 2009.

The authors declare that there are no conflicts of interest.