

Head-to-head comparison of accuracy of a rapid point-of-care HIV test with oral versus whole-blood specimens: a systematic review and meta-analysis



Nitika Pant Pai, Bhairavi Balram, Sushmita Shivkumar, Jorge Luis Martinez-Cajas, Christiane Claessens, Gilles Lambert, Rosanna W Peeling, Lawrence Joseph

Summary

Background The focus on prevention strategies aimed at curbing the HIV epidemic is growing, and therefore screening for HIV has again taken centre stage. Our aim was to establish whether a convenient, non-invasive, HIV test that uses oral fluid was accurate by comparison with the same test with blood-based specimens.

Methods We did a systematic review and meta-analysis to compare the diagnostic accuracy of a rapid HIV-antibody-based point-of-care test (Oraquick advance rapid HIV-1/2, OraSure Technologies Inc, PA, USA) when used with oral versus blood-based specimens in adults. We searched five databases of published work and databases of five key HIV conferences. Studies we deemed eligible were those focused on adults at risk of HIV; we excluded studies in children, in co-infected populations, with self-reported inferior reference standards, and with incomplete reporting of key data items. We assessed the diagnostic accuracy of testing with oral and blood-based specimens with bivariate regression analysis. We computed positive predictive values (PPVs) in high-prevalence and low-prevalence settings with Bayesian methods.

Findings In a direct head-to-head comparison of studies, we identified a pooled sensitivity about 2% lower in oral (98.03%, 95% CI 95.85–99.08) than in blood-based specimens (99.68%, 97.31–99.96), but similar specificity (oral 99.74%, 99.47–99.88; blood 99.91%, 99.84–99.95). Negative likelihood ratios were small and similar (oral 0.019, 0.009–0.040; blood 0.003, 0.001–0.034), but positive likelihood ratios differed (oral 383.37, 183.87–799.31; blood 1105.16, 633.14–2004.37). Although in high-prevalence settings PPVs were similar (oral 98.65%, 95% credible interval 85.71–99.94; blood 98.50, 93.10–99.79), in low-prevalence settings PPVs were lower for oral (88.55%, 77.31–95.87) than blood (97.65%, 95.48–99.09) specimens.

Interpretation Although Oraquick had a high PPV in high-prevalence settings in oral specimens, the slightly lower sensitivity and PPV in low-prevalence settings in oral specimens should be carefully reviewed when planning worldwide expanded initiatives with this popular test.

Funding Canadian Institutes for Health Research (CIHR KRS 102067).

Introduction

In 2004, a rapid HIV-antibody-based point-of-care test (Oraquick advance rapid HIV-1/2, OraSure Technologies Inc, PA, USA), initially approved for finger-stick, whole-blood, and plasma specimens, was approved by the US Food and Drug Administration (FDA) as a Clinical Laboratory Improvement Amendments waived test for use with specimens of oral mucosal transudate. Since 2006, with the widespread expansion of HIV testing in the USA, and with the possible expansion of home-based and new supervised self-testing initiatives in sub-Saharan Africa, this HIV test has become one of the most popular point-of-care tests based on oral specimens.^{1–3} It is more acceptable to patients because of its non-invasive and pain-free specimen collection and its rapid turnaround time.^{4–6} In Kenya and Uganda, an increased acceptance and preference for this test has helped improve the uptake of home-based HIV-testing initiatives.^{7,8} The Kenyan Government also announced an expansion of bold and controversial self-testing initiatives for HIV, and is reviewing the possible approval of oral tests.

Self-testing initiatives are also relevant for southern Africa, a region that has remained the epidemiological locus of the epidemic; countries such as Botswana, Lesotho, Mozambique, South Africa, Swaziland, Zambia, and Zimbabwe are focused on scaling up alternative HIV-screening programmes.

Oraquick is also being considered for potential use as an over-the-counter test in the USA and in many sub-Saharan countries. This move might revolutionise HIV testing by offering a proactive testing option to people who, because of stigma, do not wish to attend public health centres for testing. Hopefully, offering a confidential testing option will bring an end to the stigmatisation associated with HIV testing.⁹ Although performance data are available on this test from the USA, there has not been a review of its worldwide accuracy. With optimistic developments in HIV aimed at eradicating infection, worldwide expansion of HIV-testing programmes has taken centre stage because testing is the cornerstone of care and treatment.¹⁰ With self-testing initiatives imminent, programme planners

Published Online

January 24, 2012

DOI:10.1016/S1473-

3099(11)70368-1

See Online/Comment

DOI:10.1016/S1473-

3099(12)70003-8

Department of Medicine, McGill University, Division of Clinical Epidemiology and Infectious Diseases, McGill University Health Center, Montreal, QC, Canada

(N Pant Pai MD, B Balram BSc, S Shivkumar MSc); Division of Infectious Diseases, Department of Medicine, Queen's University, Kingston, ON, Canada

(J L Martinez-Cajas MD); Institut national de santé publique du Québec, Montreal, QC, Canada (C Claessens PhD,

G Lambert MD); Direction de santé publique de l'agence de la santé et des services sociaux de Montréal, Montreal, QC, Canada (G Lambert);

Department of Clinical Research, London School of Hygiene and Tropical Medicine, London, UK

(Prof R W Peeling PhD); and Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada (Prof L Joseph PhD)

Correspondence to:

Dr Nitika Pant Pai, Division of Clinical Epidemiology, Department of Medicine, Royal Victoria Hospital, McGill University Health Centre, V Building, 687 Pine Avenue West, Montreal, QC H3A1A1, Canada

nitika.pai@mcgill.ca

and policy makers are keen to know the relative accuracy and performance of Oraquick in oral versus blood specimens, to decide on the optimum testing algorithm.

So far, worldwide comparative data on this test have not been critically synthesised and the effect of prevalence on test accuracy has not been reviewed. This exploration is important because countries with low prevalence for HIV might consider the possible expansion of HIV-screening initiatives in the future. With a view to generating the evidence base for policy recommendations, we aimed to review worldwide evidence of this popular point-of-care test.

Methods

Search strategy and selection criteria

In accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines¹¹ we undertook a systematic review and meta-analysis to compare the diagnostic accuracy of a rapid HIV-antibody-based point-of-care test (Oraquick advance rapid HIV-1/2) when used with oral versus blood-based specimens in adults. We searched the Cumulative Index to Nursing and Allied Health Literature, Medline, Embase, BIOSIS, and Web of Science between Jan 1, 2000, and June 1, 2011. We also searched databases from key HIV conferences: International AIDS Society, Conference on Retroviruses and Opportunistic Infections, Interscience Conference on Antimicrobial Agents and Chemotherapy, Canadian Association for HIV/AIDS Research, and International Society for Sexually Transmitted Diseases. We searched bibliographies of primary studies and review articles and contacted authors for additional data. We used abstracts and brief reports when full-text articles were not available, if they contained sufficient data.¹²

To search Medline we used the string “#1 (HIV [MeSH] OR Acquired Immunodeficiency Syndrome[MeSH] OR ‘HIV Antigens’ [ti], OR ‘HIV Antibodies’[ti]), AND #2 (‘salivary’[ti] OR ‘saliva’ [ti] OR ‘blood’ [ti] OR ‘rapid’ [ti] OR ‘oral mucosal transudate’ [ti] OR ‘test’ [ti]), AND #3 (‘sensitivity’ [ti] OR ‘specificity’ [ti] OR ‘diagnostic accuracy’ [ti]) OR Oraquick[ti]”. Two reviewers (BB and SS) independently searched databases with the same search string and identified citations; a third reviewer (NPP) was consulted to resolve discrepancies.

Our review was focused on adult populations at risk for HIV; we excluded studies in children, in co-infected populations, with self-reported inferior reference standards, and with incomplete reporting of key data items.^{13–18} We also excluded editorials, perspectives, opinion pieces, manufacturer reports, and studies in other specimens. Our primary objective was to do a head-to-head comparison of the diagnostic accuracy of the test in question in oral and blood-based (finger-stick, serum, whole-blood) specimens with meta-analytic techniques. Our secondary objective was to explore the variations in positive predictive values (PPVs), with the varying prevalence recorded in studies in worldwide

settings, with a hierarchical Bayesian meta-analytic model. Lastly, we synthesised data narratively on false reactive test results happening worldwide, with a critique of data quality.

Data extraction

We used a prepiloted data abstraction form with variables such as study setting, study objectives, study populations, sample size, index test, reference standard, sensitivity, specificity, and raw cell values (true positive, false positive, false negative, true negative). Two reviewers (BB and SS) did the data abstraction and quality critique independently; disagreements were resolved by consensus with the third reviewer (NPP).

We classified reference standards as perfect or imperfect, in accordance with the guidelines of the US Centers for Disease Control and Prevention (CDC) and WHO.^{19–22} A perfect reference standard referred to one of four combinations of confirmatory testing algorithms for positive tests: dual ELISA plus whole blood, dual ELISA plus immunofluorescence assay, ELISA plus whole blood, or whole blood or immunofluorescence assay. We labelled all other combinations—ie, ELISA alone, dual ELISA—as imperfect.

Statistical analysis

For our first objective, focused on diagnostic study, we abstracted data from primary studies to obtain the four cell values of a diagnostic two-by-two table and recalculated sensitivity and specificity estimates for each study. We also visually assessed heterogeneity between studies through forest plots and also with summary receiver operating characteristic (SROC) curves with Meta-Disc software (version 1.5).²³ We plotted hierarchical summary receiver operating characteristic (HSROC) curves with STATA/IC (version 10.0). For our second objective, PPVs, we abstracted data focused on true and false positives, and recalculated PPVs with Bayesian analyses. Sensitivity and specificity estimates tend to be correlated and vary according to thresholds. HSROC curves represent summary plots of the sensitivity and specificity from the HSROC meta-analyses, with 95% joint intervals in two-dimensional space. They provide information on the overall performance of a test across different thresholds. The closer the curve is to the upper left-hand corner of the plot (sensitivity and specificity are both 100%), the better the performance of the test.²⁴ Further interpretation and details of the methods are available elsewhere.^{12,24}

For the assessment of diagnostic accuracy in oral mucosal transudate and whole blood with bivariate regression analysis, we judged that simple pooling (ie, weighted average) of sensitivity and specificity were inadequate when measures (sensitivity and specificity) were correlated. Therefore, for meta-analysis we did a random-effects bivariate regression analysis, which takes this correlation into account, and reported pooled

accuracies with 95% CIs.^{25,26} We did bivariate regression analysis in STATA/IC.^{27,28}

To explore heterogeneity we created three subgroups of studies: studies reporting head-to-head comparisons of accuracy with specimens of oral mucosal transudate and whole blood, studies reporting on specimens of oral mucosal transudate alone, and studies with specimens of whole blood alone. Hierarchically, we gave greater importance to the first subgroup because it contained studies that undertook a head-to-head comparison of samples within the same study, which removes confounding.

For our meta-analyses, we used data points obtained from individual studies: a complete set of raw cell values (ie, true positive, false positive, false negative, and true negative). A few studies reported several assessments by centres or by specimens, thus contributing to several data points.

With data on true and false positives from each study, we computed PPVs separately for specimens of oral mucosal transudate and whole blood and explored the variability of the PPV within specimen groups in low-prevalence and high-prevalence settings. We subclassified the PPVs by setting by carefully reviewing implementation research data. We defined the low-prevalence setting on the basis of estimates of seropositivity from each study and set at a conservative prevalence of disease in the study sample of less than or equal to 1%. Populations in this group included outpatients from general clinics and general population-based surveys. We defined the high-prevalence setting as greater than 1% prevalence of disease in the study sample. Populations in this group included intravenous drug users, sex workers, those who attended clinics for sexually transmitted diseases, men who have sex with men, incarcerated populations, and pregnant women. We included only studies with complete data and those done in real-life settings with cross-sectional designs, surveys, or trials. We excluded case-control studies on serum panels done in laboratories and abstracts with incomplete data (webappendix).^{29–36}

To combine PPV estimates across studies, we used a hierarchical logistic meta-analytic model. At the first level of this model, we assumed that the PPV from each study accorded with a binomial model with PPV parameters specific to each study, with the number of truly positive participants as the numerator, and the total number of positive tests as the denominator.

We assumed the logit of the binomial PPV parameters accorded with a normal density across studies, with mean representing the overall PPV across studies (on the logit scale), and the SD representing between-study variability in PPV. We completed the model by placing very wide non-informative priors on both the mean and SD of the normal density, so that inferences would be based almost entirely on the data. These numbers were directly taken from the two-by-two tables of data from each study, so that Bayes's theorem was not necessary to

derive PPV values. However, we ran three analyses to account for the effect of prevalence on the PPVs. The first analysis combined data from all studies, irrespective of prevalence. The second two analyses estimated distinct PPV values, separating studies into high and low prevalence, with a cutoff value of 1%. Being the Bayesian analogue of CIs, we report credible intervals from these Bayesian analyses. We did PPV meta-analyses with WinBUGS (version 1.4.3).

We tabulated data from studies on false reactive test results in chronological order. Although this has been referred to by previous studies, especially in the USA, we decided to include the data to provide a holistic view of the worldwide performance of Oraquick (webappendix).

Two reviewers (BB and SS) also independently rated the quality of studies, with disagreements settled by the third reviewer (NPP). We scored each item in the 14-item Quality Assessment tool for Diagnostic Accuracy Studies (QUADAS) checklist as yes, no, or unclear, and presented the results as a proportion (webappendix).^{12,37}

Role of the funding source

The sponsor of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all

See Online for webappendix

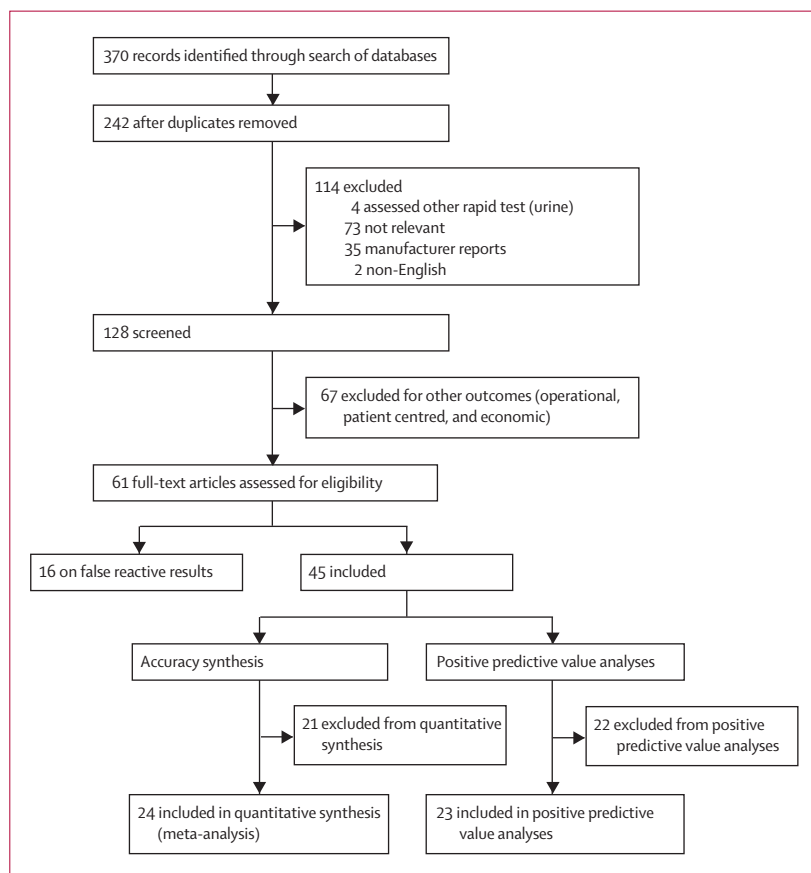


Figure 1: Study selection

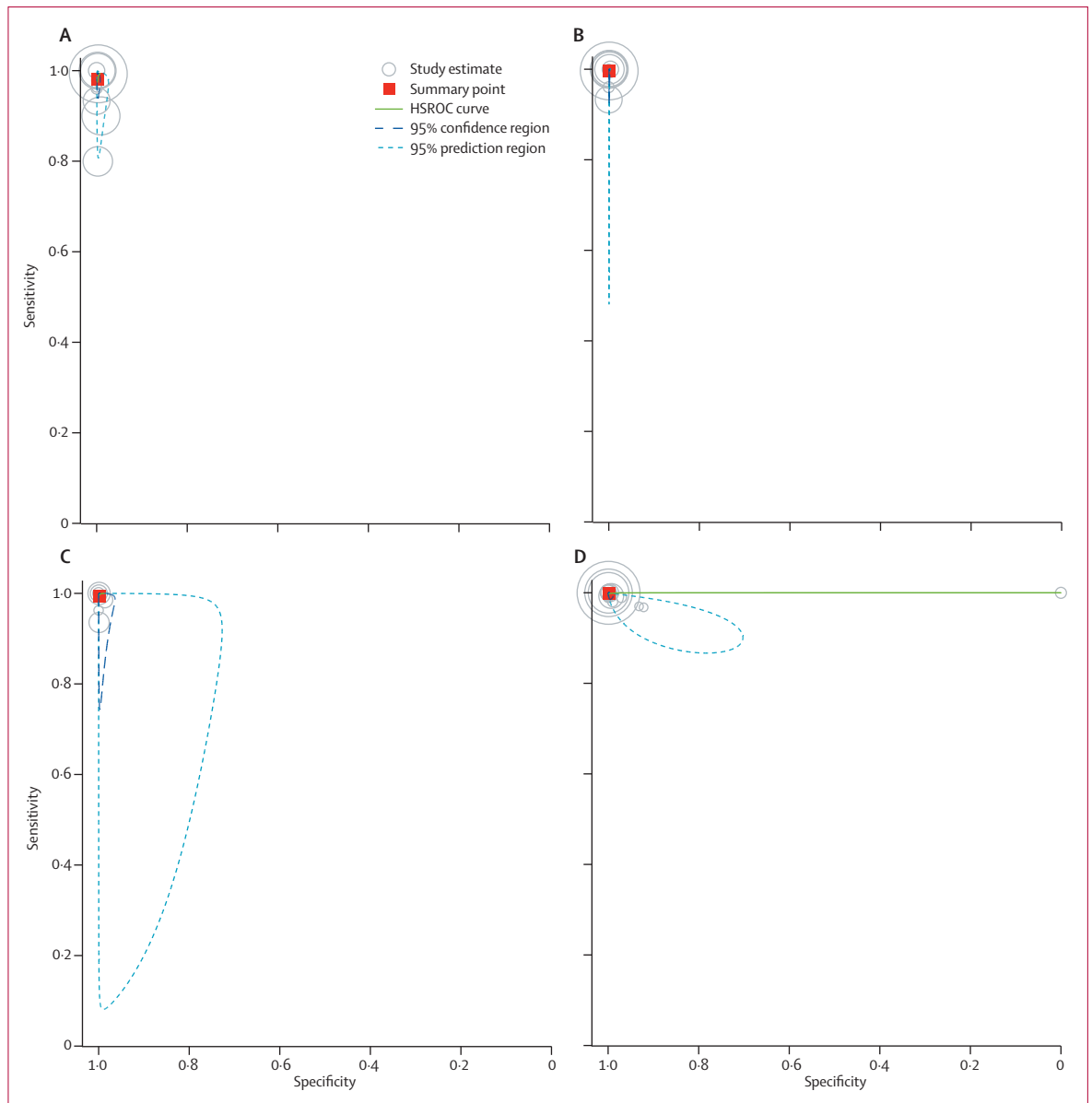


Figure 2: HSROC curves for each subgroup

Curves for studies with oral mucosal transudate within-study comparisons (A), finger-stick blood within-study comparisons (B), oral mucosal transudate samples only (C), and finger-stick blood samples only (D). HSROC=hierarchical summary receiver operating characteristic.

the data in the study and had final responsibility for the decision to submit for publication.

Results

Figure 1 shows the study selection. In our assessment of diagnostic accuracy, our pooled analyses showed seven studies in our first subgroup (studies reporting head-to-head comparisons of accuracy with specimens of oral mucosal transudate and whole blood) contributed ten data points, six studies in our second subgroup (specimens of oral mucosal transudate alone) contributed six data points, and 11 studies in our third

subgroup (specimens of whole blood alone) contributed 17 data points. Figure 2 shows our HSROC curves for each subgroup.

Our first subgroup—the main subgroup of interest—containing studies with both oral and whole-blood comparisons, provided us with the best subgroup for bivariate regression analyses. Pooled sensitivity was greater for whole-blood than oral specimens and pooled specificity was similar for each specimen (table 1). In our second subgroup, studies with no whole-blood comparators, the pooled estimates for sensitivity and specificity were similar to those for

	Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)	Log (diagnostic odds ratio)
Subgroup 1a (oral mucosal transudate within study; n=10)	98.03% (95.85–99.08)	99.74% (99.47–99.88)	383.37 (183.87–799.31)	0.019 (0.009–0.040)	9.87
Subgroup 1b (whole blood within study; n=10)	99.68% (97.31–99.96)	99.91% (99.84–99.95)	1105.16 (633.14–2004.37)	0.003 (0.001–0.034)	12.75
Subgroup 2 (oral mucosal transudate only; n=6)	99.43% (95.28–99.93)	99.86% (99.22–99.98)	721.65 (126.84–4105.76)	0.006 (0.001–0.050)	11.75
Subgroup 3 (whole blood only; n=17)	99.8% (99.07–99.93)	99.78% (99.27–99.93)	466.96 (137.42–1586.76)	0.003 (0.001–0.009)	11.78

n refers to a datapoint (one set of true positive, false positive, false negative, and true negative).

Table 1: Pooled estimates of accuracy across studies

whole-blood specimens in our first subgroup; this was also the case in our third subgroup, studies with no oral comparator.

The webappendix contains details of the studies we used in our assessment of PPV. Two studies from the CDC^{3,38} reported multiple data entries with study conduct and data collection at various sites. By use of Bayesian hierarchical meta-analytic models, we obtained estimates of PPVs for whole-blood and oral specimens in high-prevalence and low-prevalence settings. Point estimates and 95% credible intervals for PPV provided similar estimates for blood and oral specimens in high-prevalence settings (table 2). By contrast, in low-prevalence settings, PPV estimates were higher for blood than oral specimens.

We narratively synthesised data on 16 false reactive results (webappendix). We assessed risk of bias for each study. Specifically, across individual studies, we noted a pattern of incomplete reporting of test conduct, including use of separate reference standards for positives and negatives and use of a convenience sample of participants, hence we identified partial verification, sampling, selection, and detection bias biases (webappendix).

Discussion

In our first subgroup, which included studies with head-to-head comparisons of oral mucosal transudate and finger-stick specimens, the pooled sensitivity of the test in oral specimens was lower than the test's sensitivity in finger-stick specimens, a difference of about 2%. However, the specificity estimates were similar for both specimens. We give greater prominence to this comparison because within-study comparisons reduce confounding present in other subgroups because of different specimens, reference standards, settings, and devices. Six (86%) of seven studies in our first subgroup used what we defined as perfect reference standards, assessing only one device in two specimens (ie, oral mucosal transudate and finger-stick blood), forming an ideal group for within-study comparisons.

By comparing the pooled estimates from our analyses with the manufacturer's claims (sensitivity 99.3%, 95% CI 98.4–99.7; specificity 99.8%, 99.60–99.89), only

	Positive predictive value (95% credible interval)
Blood group (n=32)	
High-risk populations (n=10)	98.50% (93.10–99.79)
Low-risk populations (n=22)	97.65% (95.48–99.09)
Overall	98.03% (96.38–99.08)
Oral mucosal transudate group (n=31)	
High-risk populations (n=11)	98.65% (85.71–99.94)
Low-risk populations (n=20)	88.55% (77.31–95.87)
Overall	94.88% (87.66–98.4)

n refers to one set of true positives and false positives.

Table 2: Pooled estimates of positive predictive value

the pooled specificity estimates from our study came close to those quoted by the manufacturer. Discrepancy in sensitivity estimates from the manufacturer's estimates could be because the assessments were done in carefully controlled laboratory settings of serum panels. Also, study settings, study designs, populations, prevalence, and variable quality control procedures might affect the diagnostic performance of a test in field assessments. This difference in performance is also referred to as the optimism bias.³⁹

To put our findings into context, the lower sensitivity of the test in oral mucosal transudate compared with blood specimens is probably because of a lower quantity of HIV antibodies in oral mucosal transudate than in whole blood. The titre of HIV antibodies is also low in acute HIV infection before seroconversion, hence the increased possibility that oral testing might miss more acute HIV infections than tests with blood specimens because of its lower sensitivity. Although a very high or perfect (100%) sensitivity is desirable, it is difficult to achieve. Therefore, HIV public health programmes should emphasise this fact before recommending the oral test as a first-line screening test to detect early HIV infection in settings with low HIV prevalence. The inability of the test to identify infection during the window period must also be emphasised. Nucleic acid amplification testing and antigen antibody combination rapid tests identify infections missed by antibody-based tests.^{40,41} Therefore, for

self-testing initiatives, if the self perception of risk of a potential test taker is high, or if they suspect recent exposure to HIV and are within the window period, but their self-test result shows up as negative with an antibody-based self-test, they should be actively encouraged to seek further confirmatory testing with advanced tests immediately at a referral centre of choice.

In our meta-analysis, it is important to understand that the inherent performance characteristics of the test itself remain unchanged. What the data showed are the variations in the performance characteristics due to the amount of HIV antibodies present in the type of specimen used for testing.

Two reviews on CDC data have been published—one a comparative post-marketing assessment³ and the other a laboratory assessment.³⁸ The overall performance of the test in oral mucosal transudate was slightly lower than in specimens of whole blood.³ In a comparative study, all FDA approved blood-based point-of-care tests³⁸ assessed in the laboratory were more than 99% accurate. By comparison, only the specificity estimates in our meta-analyses were fairly close to the laboratory assessment relating to variations from implementation research data.³⁸

As our second objective we assessed the effect of surrogate prevalence estimates of HIV obtained with seropositivity estimates from each study, and their effect on PPVs. In this analysis, we noted that although the performance of the test in blood and oral specimens was similar in settings of high prevalence, the lower end of the 95% credible intervals was slightly lower for oral than for blood specimens (85.71% vs 93.10%; table 2). This variability is important to keep in mind when rolling out oral tests for expanded HIV-testing initiatives. Further, in low-prevalence settings, the test was inferior in oral compared with blood specimens. For this analysis we excluded assessments in laboratory settings and case-control designs, and focused solely on implementation research data that related to real-life settings. Because the PPV of a test is a function of the prevalence of the disease in the population, lower PPV is attributable to a large number of false positive results compared with true positive results. Subsequently, the large variability in PPV in low-prevalence settings where oral specimens were assessed implied the possibility of missing detection of new infections in settings of low prevalence and in populations at low risk of HIV acquisition. These data corroborated the data on pooled accuracy (table 1), where negative likelihood ratios were small and similar, but positive likelihood ratios differed.

To put this in context, although the oral test is popular because of its convenience and ease of specimen collection, compared with the blood-based test, the use of a single oral test in low-prevalence settings could lead to a higher number of false positives than blood-based testing. This problem could be compounded in national screening programmes and needs to be considered in the widespread

implementation of HIV testing, including home-based testing, self-testing, or over-the-counter testing initiatives, in all low-prevalence settings. Educating potential test takers on the possibility of false negatives and in those suspecting recent exposure or where clinical suspicion of positivity could be high is extremely important. In such situations, an adequate emphasis on seeking a repeat test for HIV (preferably with p24 antigen-based ELISA assays) to identify infections missed by initial screening with the oral test and optimising downstream confirmatory rapid tests (ie, nucleic acid amplification testing or western blot) will be pertinent. Therefore, because an HIV diagnosis has major implications, in initiatives such as self-testing, because Oraquick is a screening test, information on the importance of confirmatory testing must be built in or emphasised for a positive test, irrespective of specimen type. This confirmation is especially important in a low-prevalence or low-risk population, such as in pregnant women in most worldwide settings.

Our analysis has a few caveats that must be considered. Predictive values are not intrinsic attributes of a diagnostic test and are highly dependent on the prevalence of target disease. Further, a meta-analysis is used to estimate the group mean under the assumption that samples in individual studies were taken from the same population, when heterogeneity is not excessive as is evident in our meta-analysis, where homogenous subgroups were created to assess accuracy based on specimens. Furthermore, Oraquick is a diagnostic device and its performance varies with host response to HIV. Substantial biological variations in host responses as well as immunological responses take time to develop, hence the window period to allow for seroconversion. Although the test's sensitivity seems to be lower with oral versus blood specimens, both estimates obtained in our meta-analysis were at the extreme upper end of the range and there is a great deal of overlap in CIs. Hence, we could argue that the robustness of the difference is uncertain and might be affected by the results of one or two studies. The clinical significance of this difference might also be overshadowed by intrinsic variability in host status and time of testing relative to exposure—something we cannot rule out with our present analysis. Most data in our meta-analysis were reported from a high-income setting like the USA, whereas the rest of the data were from well controlled studies in developing settings. These data might not be representative of routine services in less developed countries. Finally, our review focused on Oraquick, the only FDA approved Clinical Laboratory Improvement Amendments waived test with enough worldwide data for a comparative meta-analyses. The only other test available in an oral format is Aware HIV-1/2 OMT (Calypse Biomedical Corporation, Portland, OR, USA); an over-the-counter non-FDA approved version is available on the market. This test has restricted blood versus oral comparative and independently assessed worldwide data, hence our focus on Oraquick.

No test is perfect in that there are false reactive results with almost every test, but their occurrences have been diligently recorded for Oraquick, with most evidence from the USA.

In our analysis of false positive results, we noted several responsible factors (webappendix): errors in test performance and conduct of test (ie, inaccurate specimen collection, gum swabbing more than once),³ errors in the interpretation of results (interpreting weakly reactive lines) or indeterminate test results were a direct effect of suboptimum training of counsellors, and lapses in quality assurance. Also, a cluster effect was reported from New York City (NY, USA), for reasons that were never established.⁴² Of 138 581 oral tests, about 1720 initial reactive tests were further screened by finger-stick tests and a further 353 were diagnosed as false positives.⁴² To prevent these false positives, the CDC recommended adding to the initial positive oral test a rapid finger-stick test of equal or higher accuracy, in a parallel algorithm; this algorithm could be considered for self-testing initiatives that aim to use only one screening test.⁴²

Further, a drop in test performance with kits nearing their expiration date (<1 month) was noted; this could be avoided by extending their viability period.² To sum up, these facts need to be emphasised in countries with less stringent quality control measures and where devices are used beyond their expiration dates.

In our analysis of false negative test results, most were related to the weakness of the test itself: the lack of antigen prevents the identification of an undiagnosed HIV infection.³¹ In the context of self-testing initiatives, their recurrence could be minimised by adding to the confirmatory algorithm an HIV RNA test with a shorter window period of detection in developed settings^{43,44} and cheaper antigen-antibody combination point-of-care tests, or ELISA with p24 antigens in developing settings.⁴³ Two studies on false reactive results used imperfect reference standards that yielded indeterminate results—it is important to emphasise use of the best reference standards for test assessments and use.^{43,44}

Of the 24 studies in our diagnostic meta-analysis group, 14 (60%) used a perfect reference standard as defined by CDC guidelines. We did not have enough power to explore the role of use of variable reference standards in subgroups defined by specimens.

In our quality critique of studies (webappendix), we identified that most studies were of average quality, if we weighted all items equally on the QUADAS scale, although studies in our first subgroup were of high quality.

As discussed, our meta-analysis has a potential for biases: detection, partial verification, and publication bias. Because of restricted data in each subgroup, we could not explore the role of study designs, treated versus untreated HIV infection, and reference standards in diagnostic accuracy assessments. Additionally, because of a lack of data on true-negative and false-negative values

required for negative predictive value calculations, we were unable to explore changes in negative predictive value and PPV with changes in prevalence or apply the estimated likelihood ratios and sensitivity and specificity to the full range of prevalence and plot them in a suitable graph. Lastly, the lack of independently assessed comparative worldwide data on Aware HIV-1/2 OMT prevented us from assessing it in our meta-analyses.

In this first Bayesian comparative meta-analysis of worldwide diagnostic performance data, we conclude that oral Oraquick had lower sensitivity but similar specificity to Oraquick with whole-blood specimens. Although we identified high PPVs for both oral and blood specimens in high-prevalence settings, we obtained a low PPV with oral specimens for low-prevalence settings.

Contributors

NPP, LJ, RWP, JLM-C, GL, BB, and SS conceived and designed the experiments. NPP, BB, SS, LJ, RWP, CC, GL, and JLM-C did the experiments. NPP, LJ, SS, RWP, and BB analysed the data. NPP, GL, RWP, BB, and SS contributed reagents, materials, or analysis equipment. NPP, BB, SS, LJ, JLM-C, GL, and RWP wrote the report.

Conflicts of interest

We declare that we have no conflicts of interest.

Acknowledgments

NPP acknowledges the support of the CIHR New Investigator Award 2010, and CIHR Operating grant CIHR HBF 103210.

References

- 1 CLIA database. CLIA record BP010047. Rockville, MD: US Food and Drug Administration, 2008.
- 2 Facente SN, Dowling T, Vittinghoff E, Sykes DL, Colfax GN. False positive rate of rapid oral fluid HIV tests increases as kits near expiration date. *PLoS One* 2009; 4: e8217.
- 3 Wesolowski LG, MacKellar DA, Facente SN, et al. Post-marketing surveillance of OraQuick whole blood and oral fluid rapid HIV testing. *AIDS* 2006; 20: 1661–66.
- 4 Pant Pai N, Joshi R, Dogra S, et al. Evaluation of diagnostic accuracy, feasibility and client preference for rapid oral fluid-based diagnosis of HIV infection in rural India. *PLoS One* 2007; 2: e367.
- 5 Pai NP, Barick R, Tulsy JP, et al. Impact of round-the-clock, rapid oral fluid HIV testing of women in labor in rural India. *PLoS Med* 2008; 5: e92.
- 6 Hutchinson AB, Branson BM, Kim A, Farnham PG. A meta-analysis of the effectiveness of alternative HIV counseling and testing methods to increase knowledge of HIV status. *AIDS* 2006; 20: 1597–604.
- 7 Wolff B, Nyanzi B, Katongole G, Ssesanga D, Ruberantwari A, Whitworth J. Evaluation of a home-based voluntary counselling and testing intervention in rural Uganda. *Health Policy Plan* 2005; 20: 109–16.
- 8 Negin J, Wariero J, Mutuo P, Jan S, Pronyk P. Feasibility, acceptability and cost of home-based HIV testing in rural Kenya. *Trop Med Int Health* 2009; 14: 849–55.
- 9 Pascoe SJ, Langhaug LF, Mudzori J, Burke E, Hayes R, Cowan FM. Field evaluation of diagnostic accuracy of an oral fluid rapid test for HIV, tested at point-of-service sites in rural Zimbabwe. *AIDS Patient Care STDS* 2009; 23: 571–76.
- 10 Walensky RP, Bassett IV. HIV self-testing and the missing linkage. *PLoS Med* 2011; 8: e1001101.
- 11 Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; 6: e1000097.
- 12 Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008; 149: 889–97.
- 13 Grandin W, Dev AV, Latha A, et al. Detection of human immunodeficiency virus infection in the sputum of tuberculosis patients in South India. *Int J Tuberc Lung Dis* 2010; 14: 1288–94.

- 14 Lee SR, Yearwood GD, Guillon GB, et al. Evaluation of a rapid, point-of-care test device for the diagnosis of hepatitis C infection. *J Clin Virol* 2010; **48**: 15–17.
- 15 Lejon V, Ngoyi DM, Ilunga M, et al. Low specificities of HIV diagnostic tests caused by *Trypanosoma brucei gambiense* sleeping sickness. *J Clin Microbiol* 2010; **48**: 2836–39.
- 16 Mao A, Obi-Gwacham C, Hill T, Ryan C, Hendricks A, McKellar M. Acute dengue fever causes false-positive reactivity in Oraquick rapid HIV-1/2 antibody test. *J Acquir Immune Defic Syndr* 2010; **55**: 641.
- 17 Sherman GG, Jones SA. Oral fluid human immunodeficiency virus tests: improved access to diagnosis for infants in poorly resourced prevention of mother to child transmission programs. *Pediatr Infect Dis J* 2005; **24**: 253–56.
- 18 Sherman GG, Lilian RR, Coovadia AH. Oral fluid tests for screening of human immunodeficiency virus-exposed infants. *Pediatr Infect Dis J* 2010; **29**: 169–72.
- 19 CDC. HIV testing implementation guidance for correctional settings. <http://www.cdc.gov/hiv/topics/testing/resources/guidelines/correctional-settings/> (accessed Dec 11, 2011).
- 20 Branson BM, Handsfield HH, Lampe MA, et al. Revised recommendations for HIV testing of adults, adolescents, and pregnant women in health-care settings. *MMWR Recomm Rep* 2006; **55**: 1–17.
- 21 UNAIDS/WHO. Report 11: Operational characteristics of commercially available assays to determine antibodies to HIV-1 and/or HIV-2 in human sera. Geneva: UNAIDS/WHO, 1999.
- 22 UNAIDS. HIV testing methods: UNAIDS technical update. Geneva: UNAIDS, 1997.
- 23 Zamora J, Abaira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* 2006; **6**: 31.
- 24 Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001; **20**: 2865–84.
- 25 Gatsonis C PP. Meta-analysis of diagnostic and screening test accuracy evaluations: methodologic primer. *AJR Am J Roentgenol* 2006; **187**: 271–81.
- 26 Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; **3**: 25.
- 27 Harbord R, Whiting P. Metandi: meta-analysis of diagnostic accuracy using generalized mixed models. *Stata J* 2009; **9**: 211–29.
- 28 Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003; **56**: 1129–35.
- 29 Granade TC, Parekh BS, Phillips SK, McDougal JS. Performance of the Oraquick and Hema-Strip rapid HIV antibody detection assays by non-laboratorians. *J Clin Virol* 2004; **30**: 229–32.
- 30 Barnaba AKA, Rayfield M, Fridlund C, et al. Evaluation of wholeblood HIV rapid test kits for use in VCT centres in Malawi. 15th International AIDS Conference; Bangkok, Thailand; July 11–16, 2004. MoPeB3088.
- 31 O'Connell RJ, Merritt TM, Malia JA, et al. Performance of the Oraquick rapid antibody test for diagnosis of human immunodeficiency virus type 1 infection in patients with various levels of exposure to highly active antiretroviral therapy. *J Clin Microbiol* 2003; **41**: 2153–55.
- 32 Arana FEA, Hernandez DBH, Aguilar JSA, et al. Field evaluation of rapid antibody test for detecting and confirming HIV infection in Guatemala—phase I and II. 15th International AIDS Conference; Bangkok, Thailand; July 11–16, 2004. B10467.
- 33 Holguin A, Gutierrez M, Portocarrero N, Rivas P, Baquero M. Performance of Oraquick Advance Rapid HIV-1/2 Antibody Test for detection of antibodies in oral fluid and serum/plasma in HIV-1+ subjects carrying different HIV-1 subtypes and recombinant variants. *J Clin Virol* 2009; **45**: 150–52.
- 34 Chen HT, Liang S, Liao Q, et al. HIV voluntary counseling and testing among injection drug users in south China: a study of a non-government organization based program. *AIDS Behav* 2007; **11**: 778–88.
- 35 Piwowar-Manning EM, Tustin NB, Sikateyo P, et al. Validation of rapid HIV antibody tests in 5 African countries. *J Int Assoc Physicians AIDS Care (Chic)* 2010; **9**: 170–02.
- 36 Ongarbaev AB, Avazova DE, Mustafaeva EM, et al. Detection of antibodies to HIV virus in saliva by express-tests in Uzbekistan. *Klin Lab Diagn* 2004; **8**: 40–42 (in Russian).
- 37 Whiting P, Rutjes A, Reitsma J, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; **3**: 25.
- 38 Delaney KP, Branson BM, Uniyal A, et al. Evaluation of the performance characteristics of 6 rapid HIV antibody tests. *Clin Infect Dis* 2011; **52**: 257–63.
- 39 Pai M, Minion J. Optimism bias, inflated accuracy estimates, and contradicted findings in TB diagnostic research. Invited seminar presentation Department of Epidemiology, Biostatistics and Occupational Health; Montreal, QC, Canada; Oct 5, 2009. http://www.teachepi.org/documents/courses/tbdiagrx/day1/Pai_Optimism%20bias.pdf (accessed Dec 19, 2011).
- 40 Patel P, Mackellar D, Simmons P, et al. Detecting acute human immunodeficiency virus infection using 3 different screening immunoassays and nucleic acid amplification testing for human immunodeficiency virus RNA, 2006–2008. *Arch Intern Med* 2010; **170**: 66–74.
- 41 Owen SM, Yang C, Spira T, et al. Alternative algorithms for human immunodeficiency virus infection diagnosis using tests that are licensed in the United States. *J Clin Microbiol* 2008; **46**: 1588–95.
- 42 CDC. False-positive oral fluid rapid HIV tests—New York City, 2005–2008. *MMWR Morb Mortal Wkly Rep* 2008; **57**: 660–65.
- 43 Stekler J, Swenson P, Coombs R, et al. HIV testing in a high-incidence population: is antibody testing alone good enough? *Clin Infect Dis* 2009; **49**: 444–53.
- 44 Walensky R, Arbelaez C, Reichmann W, et al. Revising expectations from rapid HIV tests in the emergency department. *Ann Intern Med* 2008; **149**: 153–60.