Interpreting the evidence-base for making recommendations on medical tests

Gopalakrishna, G.

Citation for published version (APA):
Chapter Two

Circulating antigen tests and urine reagent strips for diagnosis of active schistosomiasis in endemic areas (review)

Cochrane Database Syst Rev. 2015, 3
Abstract

Background
Point-of-care (POC) tests for diagnosing schistosomiasis include tests based on circulating antigen detection and urine reagent strip tests. If they had sufficient diagnostic accuracy they could replace conventional microscopy as they provide a quicker answer and are easier to use.

Objectives
To summarise the diagnostic accuracy of: a) urine reagent strip tests in detecting active Schistosoma haematobium infection, with microscopy as the reference standard; and b) circulating antigen tests for detecting active Schistosoma infection in geographical regions endemic for S. mansoni or S. haematobium or both, with microscopy as the reference standard.

Search methods
We searched the electronic databases MEDLINE, EMBASE, BIOSIS, MEDION, and Health Technology Assessment (HTA) without language restriction up to 30 June 2014.

Selection criteria
We included studies that used microscopy as the reference standard: for S. haematobium, microscopy of urine prepared by filtration, centrifugation, or sedimentation methods; and for S. mansoni, microscopy of stool by Kato-Katz thick smear. We included studies on participants residing in endemic areas only.

Data collection and analysis
Two review authors independently extracted data, assessed quality of the data using QUADAS-2, and performed meta-analysis where appropriate. Using the variability of test thresholds, we used the hierarchical summary receiver operating characteristic (HSROC) model for all eligible tests (except the circulating cathodic antigen (CCA) POC for S. mansoni, where the bivariate random-effects model was more appropriate). We investigated heterogeneity, and carried out indirect comparisons where data were sufficient. Results for sensitivity and specificity are presented as percentages with 95% confidence intervals (CI).
Main results
We included 90 studies; 88 from field settings in Africa. The median *S. haematobium* infection prevalence was 41% (range 1% to 89%) and 36% for *S. mansoni* (range 8% to 95%). Study design and conduct were poorly reported against current standards.

Tests for *S. haematobium*

Urine reagent test strips versus microscopy.
Compared to microscopy, the detection of microhaematuria on test strips had the highest sensitivity and specificity (sensitivity 75%, 95% CI 71% to 79%; specificity 87%, 95% CI 84% to 90%; 74 studies, 102,447 participants). For proteinuria, sensitivity was 61% and specificity was 82% (82,113 participants); and for leukocyturia, sensitivity was 58% and specificity 61% (1532 participants). However, the difference in overall test accuracy between the urine reagent strips for microhaematuria and proteinuria was not found to be different when we compared separate populations (P = 0.25), or when direct comparisons within the same individuals were performed (paired studies; P = 0.21).

When tests were evaluated against the higher quality reference standard (when multiple samples were analysed), sensitivity was marginally lower for microhaematuria (71% vs 75%) and for proteinuria (49% vs 61%). The specificity of these tests was comparable.

Antigen assay
Compared to microscopy, the CCA test showed considerable heterogeneity; meta-analytic sensitivity estimate was 39%, (95% CI 6% to 73%); specificity 78%, (95% CI 55% to 100%) (four studies, 901 participants).

Tests for *S. mansoni*
Compared to microscopy, the CCA test meta-analytic estimates for detecting *S. mansoni* at a single threshold of trace positive were: sensitivity 89% (95% CI 86% to 92%); and specificity 55% (95% CI 46% to 65%; 15 studies, 6091 participants) Against a higher quality reference standard, the sensitivity results were comparable (89% vs 88%) but specificity was higher (66% vs 55%). For the CAA test, sensitivity ranged from 47% to 94%, and specificity from 8% to 100% (four studies, 1583 participants).

Authors’ conclusions
Among the evaluated tests for *S. haematobium* infection, microhaematuria correctly detected the largest proportions of infections and non-infections identified by microscopy.
The CCA POC test for *S. mansoni* detects a very large proportion of infections identified by microscopy, but it misclassifies a large proportion of microscopy negatives as positives in endemic areas with a moderate to high prevalence of infection, possibly because the test is potentially more sensitive than microscopy.

**Plain language summary**

**How well do point-of-care tests detect Schistosoma infections in people living in endemic areas?**

Schistosomiasis, also known as bilharzia, is a parasitic disease common in the tropical and subtropics. Point-of-care tests and urine reagent strip tests are quicker and easier to use than microscopy. We estimate how well these point-of-care tests are able to detect schistosomiasis infections compared with microscopy.

We searched for studies published in any language up to 30 June 2014, and we considered the study’s risk of providing biased results.

**What do the results say?**

We included 90 studies involving almost 200,000 people, with 88 of these studies carried out in Africa in field settings. Study design and conduct were poorly reported against current expectations. Based on our statistical model, we found:

- Among the urine strips for detecting urinary schistosomiasis, the strips for detecting blood were better than those detecting protein or white cells (sensitivity and specificity for blood 75% and 87%; for protein 61% and 82%; and for white cells 58% and 61%, respectively).

- For urinary schistosomiasis, the parasite antigen test performance was worse (sensitivity, 39% and specificity, 78%) than urine strips for detecting blood.

- For intestinal schistosomiasis, the parasite antigen urine test, detected many infections identified by microscopy but wrongly labelled many uninfected people as sick (sensitivity, 89% and specificity, 55%).
What are the consequences of using these tests?

If we take 1000 people, of which 410 have urinary schistosomiasis on microscopy testing, then using the strip detecting blood in the urine would misclassify 77 uninfected people as infected, and thus may receive unnecessary treatment; and it would wrongly classify 102 infected people as uninfected, who thus may not receive treatment. If we take 1000 people, of which 360 have intestinal schistosomiasis on microscopy testing, then the antigen test would misclassify 288 uninfected people as infected. These people may be given unnecessary treatment. This test also would wrongly classify 40 infected people as uninfected who thus may not receive treatment.

Conclusion of review

For urinary schistosomiasis, the urine strip for detecting blood leads to some infected people being missed and some non-infected people being diagnosed with the condition, but is better than the protein or white cell tests. The parasite antigen test is not accurate.

For intestinal schistosomiasis, the parasite antigen urine test classifies many microscopy negative people as being infected. This finding may be explained by the low sensitivity of microscopy.

Background

Target condition being diagnosed Schistosomiasis, also known as bilharzia, is the second major parasitic disease affecting tropical and subtropical regions after malaria. It is caused by trematode worms of the genus Schistosoma (Gryseels 2012). The latest estimates show that schistosomiasis is endemic in 76 countries, with 779 million people at risk of infection and approximately 207 million people currently infected. Sub-Saharan Africa accounts for more than 90% of current cases of schistosomiasis (Engels 2002; WHO 2010; Gryseels 2012). The global burden of disease in 2004 was estimated at 13 to 15 million disability-adjusted life-years (DALYs) lost as the result of schistosomiasis (King 2010a). These estimates could be an underestimate resulting from the low sensitivity of routinely used diagnostic tests (King 2010a; King 2010b).

Five main schistosome species are known to infect man (*Schistosoma mansoni*, *Schistosoma haematobium*, *Schistosoma japonicum*, *Schistosoma intercalatum*, and *Schistosoma mekongi*), of which *S. mansoni*, *S. haematobium*, and *S. japonicum* have the greatest impact on morbidity (Gryseels 2006). The focus of this review will be on diagnosing...
infection caused by *S. mansoni* and *S. haematobium*, as they are more widespread globally and account for most infections and associated morbidity worldwide. These species cause intestinal schistosomiasis and urogenital schistosomiasis, respectively. Urogenital schistosomiasis presents with blood in urine (haematuria), proteins in urine (proteinuria), or white blood cells in urine (leukocyturia). In its chronic form, it presents with major bladder, kidney, and genital pathologies including chronic renal failure. Intestinal schistosomiasis presents with abdominal pain and in its chronic and severe forms can present with enlarged liver (hepatomegaly), abdomen distended with fluid (ascites), and liver failure.

Currently, no vaccine is available to protect against schistosomal infection (Rollinson 2009; Bethony 2011). If left untreated, schistosomal infection may result in chronic disease. The current drug of choice is praziquantel, which is cheap (costing less than USD 0.15 per treatment) and safe and causes few side effects. Praziquantel however is ineffective against the eggs and larval forms of schistosome worms (Gryseels 2012; Rollinson 2013). Mass praziquantel treatment of populations at risk of infection is now routine in many endemic areas (WHO 2010; Rollinson 2013). Reinfections rapidly occur as the result of recurrent direct contact with water bodies infected with schistosomal parasites (WHO/TDR 2006; Rollinson 2009; Rollinson 2013). No strong evidence of clinically relevant drug resistance is available (Geerts 2001; Doenhoff 2002; Fenwick 2003; Doenhoff 2009; Greenberg 2013).

However reports have described heterogeneities in egg reduction rates and in systematic non-clearers of infection after treatment with praziquantel (Black 2009; Melman 2009; Ahmed 2012). In the long run, mass treatment has limitations related to cost-effectiveness (French 2010), poor sustainability (Utzinger 2009), poor drug compliance by individuals (Guo 2005; Croce 2010), and increased drug selection pressure (Greenberg 2013). Accurate and affordable diagnostic tools are essential for providing targeted treatment and for maximizing the success of control of schistosomiasis in endemic areas; they are required for monitoring drug efficacy as well. Diagnosis of schistosomiasis can be performed directly or indirectly. Direct methods include detection of schistosome eggs in urine or stool by microscopy, detection of schistosome antigens in serum or urine samples, and detection of Schistosoma-specific DNA in urine, stool, or blood. Indirect methods include questionnaires, biochemical tests (urine reagent strips for microhaematuria/proteinuria/leukocyturia), antibody tests, ultrasonography, computed tomography (CT) scan, magnetic resonance imaging (MRI) scan, endoscopy, and cystoscopy (Feldmeier 1993; Rabello 1997; Doenhoff 2004; Bichler 2006; Gryseels 2012; Cavalcanti 2013).
Currently no gold standard is recommended for the detection of schistosomiasis. Microscopy is the most widely used test for diagnosing schistosomiasis and, although imperfect, it is commonly used as the reference standard in practice. Its sensitivity has been shown to vary with intensity of infection, prevalence of infection, sample preparation techniques, stool consistency, and circadian and day-to-day variation of egg counts in stool and/or urine (Doehring 1983; Doehring 1985a; Rabello 1992; Feldmeier 1993; Rabello 1997; van Lieshout 2000; Knopp 2008). This becomes particularly pertinent as control programmes progress and sensitivity of microscopy decreases as the result of reduced infection intensity. Repeated measurements over multiple days from multiple samples and/or multiple smears/slides taken from each sample has been shown to increase sensitivity (Knopp 2008; da Frota 2011; Siqueira 2011; Deelder 2012); however this task increases the time taken to perform the survey and therefore becomes logistically expensive (van Lieshout 2000; Legesse 2007).

**Index test(s)**

Urine reagent strips and circulating antigen tests are used as alternatives to microscopy for diagnosis of schistosomiasis. Compared with microscopy, urine reagent strips used to detect microhaematuria or proteinuria as a proxy for *S. haematobium* infection are cheap, quick, and easy to use (Mott 1985; Brooker 2009); have no technical requirements; and are less influenced by the circadian production of schistosome eggs (Murare 1987; Lengeler 1991b). Furthermore, some studies have shown that the sensitivity of these strips is higher than that of urine filtration (French 2007; Robinson 2009), and that a single test with microhaematuria strips is more sensitive than a single test with urine filtration (Taylor 1990).

These features make these strips suitable for screening of urogenital schistosomiasis in the field. However, results should be interpreted against the background of risk for schistosomiasis, as well as any other signs and symptoms that could be indicative of other diseases. Microhaematuria and proteinuria are non-specific signs that could also result from other ailments such as urogenital infection, malignancy, immune system disorders, metabolic disorders, and trauma.

Circulating antigen tests (circulating anodic antigen (CAA) and circulating cathodic antigen (CCA)) have also been evaluated as replacements formicroscopy in the diagnosis of infection due to *S. haematobium* or *S. mansoni*. These tests can differentiate between active and past infections, as the circulating antigens are probably present only when
there is active infection (Doenhoff 2004). As circulating antigens are released from living worms, antigen levels may correlate directly with parasite load, whilst microscopy does not. This may make the CCA POC test useful in monitoring the dynamics of worm burdens and clearance of worms after treatment (Cavalcanti 2013; Rollinson 2013). However, the sensitivity of these tests has been shown to vary with prevalence of disease and intensity of infection (De Jonge 1988; De Jonge 1989; van Lieshout 1992; De Clerq 1997; Stothard 2006; Ayele 2008; Obeng 2008; Midzi 2009; Colley 2013).

This review evaluates the urine CCA POC test, urine CCA and CAA enzyme-linked immunosorbent assay (ELISA), and serum CCA and CAA ELISA. The urine CCA POC test is a lateral flow assay that uses a nitrocellulose strip with a monoclonal antibody coated test line to detect the presence of Schistosoma-specific CCA antigen in urine. When urine from an infected individual flows through the strip, the antigen will bind to the test line, which becomes visible with the binding of added labelled monoclonal antibodies (van Dam 2004). Of note, the urine CCA POC test was developed based on the performance of the ELISA format (Brooker 2009). The urine CCA ELISA was found to have the best diagnostic performance, followed by the serum CAA assay for \textit{S. mansoni} (Polman 1995; van Lieshout 1995; van Lieshout 2000). Therefore, although they are not rapid tests, the accuracy measures of ELISA tests will be systematically assessed, as the summary measures obtained may guide the ongoing development of improved POC tests.

So far, a range of accuracy measures have been reported for urine reagent tests and for circulating antigen tests. Diagnostic and treatment strategies in endemic areas vary with results of these tests and depend on financial and human resource capacity.

**Clinical pathway**

Patients suspected of having active \textit{S. haematobium} or \textit{S. mansoni} infection in endemic settings.

**Prior test(s)**

Current practice in endemic settings is to use urine reagent strips as a replacement for microscopy or as a triage test (before microscopy), or circulating antigen tests as a replacement for microscopy. In line with practice in disease control programmes, we focus on the role of these tests as alternatives to microscopy. We will not consider prior testing with other tests, as this is rarely done in public health programmes.
Role of index test(s)
We are interested in the following purposes for testing.

- Reagent strips to detect microhaematuria, proteinuria, or leukocyturia as a replacement test for microscopy for *S. haematobium* infection.

- CCA point-of-care test as a replacement test for microscopy for *S. haematobium* or *S. mansoni* infection.

Alternative test(s)
Apart from the two test types mentioned above, a range of other tests can be used to screen for schistosomiasis. However, all are used in different situations and in different circumstances than the tests mentioned above.

Questionnaires have been used for the initial rapid screening for urinary schistosomiasis in high-risk communities in endemic areas (Lengeler 1991a; Feldmeier 1993; Chitsulo 1995). These questionnaires rely on self-reporting of blood in urine. Studies have shown that questionnaires demonstrate moderate to high sensitivities and specificities when used to screen individuals for urogenital schistosomiasis in high-prevalence areas but low sensitivity and specificity in low-prevalence areas (Lengeler 1991a; Lengeler 1991b; Brooker 2009). Questionnaires for intestinal schistosomiasis have been shown to be less sensitive and specific than those for urogenital schistosomiasis (WHO/TDR 2006; Brooker 2009).

Symptoms of intestinal schistosomiasis are associated with many other diseases, which often overlap in range. As co-infection is the norm rather than a rare occurrence, the questionnaires are less specific. The accuracy of questionnaires has been shown to be influenced by age and gender. When questionnaires are used repeatedly in the same area, respondents are prone to give biased answers, as they know the consequences of the answers they give. Thus, recall bias may interfere with the accuracy of the test. Consequently, relying on questionnaires may become ineffective, making this screening method unsuitable even for follow-up of patients after treatment (Ansell 1997; Guyatt 1999; Lengeler 2002).

As questionnaires are recommended mainly for initial rapid screening and not for routine screening for schistosomiasis, they will not be evaluated in this review. Serology tests
are alternative tests for the diagnosis of schistosomiasis. These tests detect antibodies against worm antigens, egg antigens (soluble egg antigens (SEAs)), or eosinophil cationic proteins (ECPs) (Reimert 1991; Feldmeier 1993; ITM 2007). Available methods include ELISA, indirect immunofluorescence assay (IFA), and indirect haemagglutination assay (IHA). Antibody tests demonstrate high sensitivity even in areas with light infection and therefore can be used in areas with low endemicity. However, these tests fall short in distinguishing current active infection from past infection, have low specificity in endemic areas because of cross-reactivity with antigens of other helminths, and often show antibody levels that remain elevated after treatment; therefore, they yield many false-positive results (Doenhoff 2004; Cavalcanti 2013). Antibody tests may have a role in checking for maintained exposure to schistosomiasis in areas that are moving towards elimination (Rollinson 2013).

The ECP test is an indirect marker of *S. haematobium* infection and related morbidity (Reimert 2000; Vennervald 2004). Other test examples include rectal biopsy (ITM 2007), cystoscopy and endoscopy, radiological methods (Bichler 2006), FLOTAC (a novel faecal egg count technique) (Knopp 2009; Glinz 2010), and molecular tests using polymerase chain reaction (PCR) (Ten Hove 2008; Oliveira 2010; Knopp 2011). However, these tests may be expensive or may require trained laboratory personnel and an elaborate laboratory infrastructure.

**Rationale**

For improved mapping to ensure effective selective (or targeted) treatment and for accurate data on treatment success with praziquantel, appropriate diagnostic tests are urgently required. When a test for diagnosing schistosomiasis is considered, a test with high sensitivity is paramount, especially when infection is being monitored within a disease control programme. False-negative results lead to missed treatment and subsequently to more advanced disease, or, if occurring after praziquantel treatment, may lead to overestimated cure rates and potentially undetected cases of praziquantel resistance and the spread of the disease. High specificity is also required, as unnecessary treatment due to false-positive results could reduce cost-effectiveness in current control programme strategies through potentially inaccurate classification of prevalence levels or in future targeted treatment control programmes (WHO/TDR 2006).

On the other hand, a test for mapping of disease (to get an estimation of disease prevalence in an endemic area) may not need sensitivity and specificity as high as those
required for monitoring of disease. There is currently no recommended gold standard for the detection of active schistosomiasis. However, because microscopy is the most commonly used test in practice and is often used as the reference test in studies, we selected it for use as the reference standard within this review to detect *S. haematobium* and *S. mansoni*. The primary concern with microscopy is the possibility of missing infected cases (because of its low and varied sensitivity), especially in areas with low intensity of infection. This means that truly infected cases may be missed and misclassified as non-infected by microscopy. Therefore when comparing an index test against microscopy, the number of false-positives (potentially true cases classified as positive by the index test and classified as negative by the reference test) may be high, and the index test may present with low specificity.

Increasing the sensitivity of microscopy by taking multiple measurements may reduce the number of true cases wrongly classified as non-infected by microscopy. An index test compared against a more sensitive reference test (microscopy with multiple measurements) may have higher specificity because the number of false-positives will be low. Our review will therefore also investigate the effect of the quality of the reference standard on the sensitivity and specificity of the index tests being evaluated.

In this case, a test considered as a replacement for microscopy should have comparable sensitivity or should be less costly, portable, faster, and easier to use or interpret, and it should be less demanding logistically. Point-of-care tests based on circulating antigen detection and biochemical urine reagent strips in particular are being included (or developed) in disease control strategies, as they are easy to use and interpret, require minimal laboratory infrastructure, are cost-effective, reduce patient waiting time and potentially therefore reduce loss to follow-up, and may have comparable or higher sensitivity to microscopy (Loubiere 2010). The results of this review may guide policy makers on appropriate diagnostic tests to use and may help identify research gaps in diagnostic testing for schistosomiasis in endemic areas.

**Objectives**

With the goals of making recommendations and informing policy makers on which tests to use and identifying research gaps, these were our primary objectives:

- To obtain summary estimates of the diagnostic accuracy of urine reagent strip tests for microhaematuria, proteinuria, and leukocyturia in detecting active *S. haematobium*
infection, with microscopy of urine as the reference standard.

- To obtain summary estimates of the diagnostic accuracy of circulating antigen tests—a urine POC circulating cathodic antigen (CCA) test, a urine and serum CCA enzyme-linked immunosorbent assay (ELISA) test, and a urine and serum circulating anodic antigen (CAA) test—for detection of active Schistosoma infection in geographical regions endemic for *S. mansoni* or *S. haematobium* or both, with microscopy as the reference standard.

- To compare the accuracy of the above index tests.

- To investigate potential sources of heterogeneity in the diagnostic accuracy of the tests listed above.

**Secondary objectives**

To investigate whether age and gender of participants, positivity thresholds, prevalence of infection, intensity of infection, quality of the reference standard, effects of praziquantel treatment, infection stage, mixed infections, and the methodological quality of included studies can explain observed heterogeneity in estimates of test accuracy.

**Methods**

**Criteria for considering studies for this review**

**Types of studies**

We included primary observational studies that compared the results of one or more of the index tests versus the reference standard. These studies could be cross-sectional in design, cohort studies, or diagnostic case-control studies with cases and controls sampled from the same patient population.

We included studies that provide participant data. Only studies in which true-positives (TPs), true-negatives (TNs), false-positives (FPs), and false-negatives (FNs) were reported or could be extracted from the data were included.

We excluded case-control studies with healthy controls, controls from non-endemic areas, or controls with alternative diagnoses (patients with diseases similar to schistosomiasis), as specificity may be overestimated (Rutjes 2005). False-positive test results may
occur when an alternative disease produces the same pathophysiological changes as the target condition. We also excluded studies that enrolled only participants with proven schistosomiasis, as sensitivity may be overestimated.

**Participants**
Participants had to be individuals residing in regions where *S. haematobium* and *S. mansoni* infections were endemic. We excluded articles that studied travelers originating from non-endemic countries, as they were typically screened with other tests such as antibody tests.

**Index tests**
We included studies that evaluated the following tests

**Urine reagent strip tests**
A urine reagent strip test is a biochemical semiquantitative test. It is regarded as an indirect indicator of *S. haematobium* infection or morbidity, as it detects microhaematuria, proteinuria, or leukocyturia (white blood cells in urine) that can develop as a consequence of schistosomal infection (Doehring 1985b; Doehring 1988). This test is cheap and easy to use for rapid screening of urinary schistosomiasis (Feldmeier 1993; Gryseels 2006; Gryseels 2012).

The results of urine reagent tests used to measure haematuria are scored as 0 (negative), trace-positive (tr), 1+ (5 to 10 erythrocytes/μL), 2++ (10 to 50 erythrocytes/μL), or 3+++ (50 to 250 erythrocytes/μL). For proteinuria, results are scored as 0 (negative), trace-positive (tr), 1+ (30 mg protein/dL), 2++ (100 mg protein/dL), or 3+++ (500 mg protein/dL) (Murare 1987).

**Antigen tests**
Antigen tests are based on detection of schistosome antigens in the serum and urine of individuals (Gryseels 2006; WHO/TDR 2006; Gryseels 2012). The main circulating antigens are adult worm gut-associated circulating antigens, and CAA and CCA are the main focus of research.

The CCA dipstick is scored according to test band reaction intensity as negative (-), trace-positive (tr), single-positive (+), doublepositive (++), and triple-positive (+++) (Stothard 2006). ELISA results are continuous, and positivity thresholds may vary. To esti-
mate the accuracy of ELISA tests, ELISA must have been evaluated against the reference standard only.

**Target conditions**
Active infection with *S. haematobium*.
Active infection with *S. mansoni*.

**Reference standards**
*S. haematobium*
For diagnosis of *S. haematobium* infection, the reference standard is microscopy of urine for examination of schistosome eggs. To increase sensitivity, urine samples can be concentrated by sedimentation, filtration, or centrifugation techniques (Gryseels 2006), or more samples can be examined (Feldmeier 1993). We therefore included studies that use all of these concentration techniques, and to estimate the effect of the quality of the reference standard, we accepted studies using microscopy on a single urine sample (lower quality reference standard) and studies performing microscopy on multiple urine samples (higher-quality reference standard).

*S. mansoni*
For diagnosis of *S. mansoni* infection, microscopic examination of schistosome eggs in stool is the reference standard. Sensitivity is increased by preparing a faecal thick smear using the Kato-Katz (KK) method (Gryseels 2006) or by examining multiple stool samples (Feldmeier 1993). To estimate the effect of the quality of the reference standard, we accepted studies using microscopy on a single stool sample (lower-quality reference standard) and studies performing microscopy on multiple stool samples (higher-quality reference standard).

It is important to note that some regions experience mixed infections of *S. haematobium* and *S. mansoni*. In such situations, microscopy of both stool and urine samples must be carried out to confirm infection.

**Search methods for identification of studies**

**Electronic searches**
We searched the electronic databases MEDLINE, EMBASE, BIOSIS, MEDION, and HTA (Health Technology Assessment). We further translated the MEDLINE search to EMBASE
and BIOSIS databases to identify additional records. To avoid missing studies, we did not use a diagnostic search filter. We performed the searches on 12 January 2012 and repeated them on 16 November 2012, 29 August 2013, and 30 June 2014.

Searching other resources
We looked through reference lists of relevant reviews and studies and websites of the World Health Organization (WHO), the Schistosomiasis Control Initiative (SCI), and the Schistosomiasis Consortium for Operational Research and Evaluation (SCORE). When possible, we contacted study authors to request extra information.

Data collection and analysis

Selection of studies
Two independent review authors first looked through titles and abstracts to identify potentially eligible studies. Full-text articles of these studies were obtained and assessed for study eligibility by two independent review authors using the predefined inclusion and exclusion criteria. Disagreements were resolved through discussion and by consultation with a third review author when necessary.

Data extraction and management
Two independent review authors extracted data onto a data extraction form. The following data were extracted.

- Study authors, publication year, and journal.
- Study design.
- Study participants-age, sex.
- Prevalence of schistosomiasis.
- Treatment status of participants with praziquantel treatment status before study or post treatment.
- Reference standard (microscopy), including number of samples per individual and exact volume of stool/urine examined.
• Index tests—urine and serum circulating antigen tests (CCA and CAA) and urine reagent strips.

• Urine reagent strips—signs measured (microhaematuria, proteinuria, leukocyturia).

• Sample preparation techniques—time of day urine/stool sample was taken, intensity of infection—egg counts in urine and stool by microscopy.

• Presence of missing or unavailable test results.

• Numbers of TPs, FNs, FPs, and FNs.

**Assessment of methodological quality**

We used the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool to assess risk of bias and concerns for applicability of the included studies (Whiting 2011). Disagreements were resolved through consensus or by consultation with a third review author. We extracted data using signalling questions and scored for risk of bias and concerns for applicability under the four main domains: participant selection, index test, reference standard, and participant flow.

**Statistical analysis and data synthesis**

**Comparisons of index test versus the reference standard**

We analyzed data for the two target conditions (*S. haematobium* and *S. mansoni*) separately. Only one included study (Ashton 2011) evaluated the ability of a test to detect *S. haematobium* and/or *S. mansoni* in an area of mixed infection. Among studies reporting sufficient data for calculating sensitivity and specificity, we plotted their sensitivity and specificity in both forest plots and receiver operating characteristic (ROC) space using the software Review Manager 5.2. We performed meta-analysis using the statistical software SAS version 9.2 for test types that had sufficient data points (four or more data points) to be pooled by the statistical models and those that did not demonstrate substantial heterogeneity in ROC space (Macaskill 2010). These tests included the reagent strip for microhaematuria, the reagent strip for proteinuria, the reagent strip for leukocyturia, the CCA POC test for *S. haematobium*, and the CCA POC test for *S. mansoni*. 
The statistical model selected to perform the overall meta-analysis depended on the variability of the positivity thresholds, as discussed below. Data for urine reagent strips and urine CCA POC tests were ordinal. These tests are typically scored as 0, trace, 1+, 2+, and 3+, or as 0, 1+, 2+, and 3+. When data from a test had multiple thresholds, we used the hierarchical summary receiver operating characteristic model (HSROC) to perform the overall meta-analysis. This model estimates the underlying ROC curve, which describes how sensitivity and specificity of the included studies trade off with each other as thresholds vary. It allows for variation in the parameters of accuracy, thresholds between studies, and the shape of the underlying ROC curve (Rutter 2001; Macaskill 2010). Because this method models sensitivity and specificity indirectly, we calculated average sensitivities and average specificities from the output of the model. When data from a test had one or a common threshold, we used the bivariate random-effects model to perform the overall meta-analysis. This method models sensitivity and specificity directly at a common threshold (Reitsma 2005; Macaskill 2010).

We included all studies in the overall meta-analysis, whether or not a positivity threshold was included. We assumed that different thresholds were used for the studies that did not report their thresholds, and we used the HSROC model to perform the overall meta-analysis. For urine reagent strips for microhaematuria and proteinuria, many studies did not report a positivity threshold (n = 41 for microhaematuria and n = 25 for proteinuria). Some studies (n = 2) provided data points at both thresholds of trace and +1.

When data points were provided at both thresholds, we selected the data point at threshold trace for the overall analysis; we selected the first stipulated positivity threshold. Leukocyturia had five overall data points, with four data points at threshold trace and one at +1. The CCA POC for S. haematobium had four overall data points, with two at threshold trace and two at +1.

All studies evaluating CCA POC for S. mansoni reported positivity thresholds; five provided data points at both thresholds trace and +1. When data points were provided at both thresholds, we selected the data point at threshold trace for the overall analysis; we selected the first stipulated positivity threshold. The overall analysis therefore contained 15 data points with threshold ≥ trace, for which we used the bivariate model for meta-analysis.

Comparisons of index tests

We compared the accuracy of the reagent strips for microhaematuria in detecting S. haematobium...
Interpreting the Evidence-base for Making Recommendations on Medical Tests

matobium versus the accuracy of the reagent strips for proteinuria. These were the only tests with sufficient data to enable comparisons between different types of tests. Tests were compared by adding the co-variante test type to the HSROC model and allowing this to have an effect on the accuracy, threshold, and shape parameters. We performed indirect comparisons and direct comparisons; in the latter, we included only studies that applied both index tests in the same individuals.

Investigations of heterogeneity

We investigated heterogeneity by examining the forest plots and statistically by including co-variates in the HSROC or bivariate model, by conducting subgroup analysis, and by performing sensitivity analysis. In the HSROC model, we investigated whether these co-variates affect the parameters of this model—accuracy, threshold, and shape—whereas in the bivariate model, we investigated whether these co-variates affect sensitivity and specificity. We did not investigate the effects of infection stage and mixed infection caused by poor reporting and insufficient data for these items. We investigated the following sources of heterogeneity: quality of the reference standard, positivity threshold, age, gender (proportion of female participation), intensity of infection, prevalence of infection, effect of praziquantel treatment, and QUADAS-2 risk of bias domains. Of these, the co-variates gender (proportion of female participation) and prevalence of infection were analyzed as a continuous co-variante. The rest were analyzed as categorical co-variates. We classified studies that used single-measurement microscopy (one stool and/or one slide or smear) and those that did not report how the reference standard was conducted as using lower-quality reference standards because single measurements are more likely to miss diseased individuals.

We assumed that studies that used multiple measurements of microscopy were likely to report this, given the relevance of this additional effort. Reference standards that used multiple urine or stool samples or multiple slides or smears were classified as higher-quality reference standards. For the age co-variante, many mixed adult/children studies did not state the proportions of adults or children. Some did not state the age of participants. As accuracy data were not provided for age subgroups in most studies, we dichotomized the age co-variante into the groups 'all ages' and 'children only’. We assumed that studies that did not state the age had included participants of all ages. Because the proportions of female and male participants were poorly reported at the test level and at the level of the 2 × 2 tables, we analyzed the co-variante of gender as a continuous variable at the study level. For this co-variante, gender indicated the proportion of female participation. We focused on females because gender may influence
accuracy estimates through factors associated with females, such as menstruation and genitourinary tract infection (Hall 1999; French 2007; Brooker 2009).

The World Health Organization (WHO) recommendations (WHO 2002) categorize intensity of infection for S. haematobium as follows: < 50 eggs/10 mL (light) and ≥ 50 eggs/10 mL (heavy) and intensity of S. mansoni as follows: 1 to 99 eggs per gram (epg) (light), 100 to 399 epg (moderate), and ≥ 400 epg (heavy). In our review, the intensity of infection was reported in different ways (arithmetic mean or range of infection, or geometric mean or range of infection, or proportions of participants with light/moderate/heavy infection) and for most included studies was not reported at all (63% and 65% for microhaematuria and proteinuria, respectively). We used the reported estimates of mean (arithmetic/geometric) or median intensity of infection to classify our studies according to WHO recommendations. We classified as unclear studies that reported only proportions of participants with light/moderate heavy infections or did not report estimates of intensity of infection.

We examined the effects of treatment with praziquantel on the sensitivity and specificity of the test type microhaematuria because it was the only test with sufficient data to investigate this. Nine studies provided data on praziquantel treatment; seven were follow-up studies with praziquantel given at variable intervals (King 1988_a (one year), NGoran 1989 (one month), Kitange 1993 (one year), Lengeler 1993 (one month), Shaw 1998 (six weeks), Magnussen 2001 (one year), French 2007 (one year)), and two indicated that praziquantel had been given before the baseline study was performed (Abdel-Wahab 1992 (two years), Bogoch 2012 (two years)). When multiple follow-up studies were performed, we selected data for the first follow-up evaluation (Shaw 1998; French 2007). However, pooling of results of all studies with varying time intervals would likely introduce a lot of heterogeneity, bias our summary estimates, and lead to overestimates of sensitivity, because studies with long time intervals were likely to have a greater number of participants reinfected compared with studies done at shorter time intervals. We opted to present estimates of sensitivity and specificity of individual studies evaluating the performance of microhaematuria post treatment in the ROC space. We added the following co-variates one by one to the HSROC model for microhaematuria and proteinuria and to the bivariate model for CCA POC for S. mansoni: quality of the reference standard, age, gender, and prevalence of infection. We then performed a subgroup analysis for the co-variates-quality of the reference standard, age, positivity threshold, and intensity of infection-for all three index tests.
Sensitivity analyses
We performed a sensitivity analysis to check the robustness of results when filtration was used as a concentration for urine microscopy for *S. haematobium*, and to estimate sensitivity and specificity for studies with low risk of bias according to the QUADAS domains, along with participant selection, participant flow, and the reference standard.

Assessment of reporting bias
We did not assess reporting bias. Methods of assessing reporting bias for diagnostic accuracy studies are still being refined. For instance, the Deeks test, a test that has been proposed for use in diagnostic accuracy studies, has low power to detect funnel plot asymmetry, especially when a lot of heterogeneity is present (Macaskill 2010). The studies included in our review showed a lot of heterogeneity; therefore assessments for reporting bias may not yield conclusive results.

Results

Results of the search
Our search yielded 17,477 hits. After the titles and abstracts were screened, 152 full texts were retrieved, and after full texts were assessed, 90 articles were deemed suitable for inclusion; 62 were excluded. One study author whom we contacted responded to our request for information, but the data submitted did not meet our eligibility criteria. No additional eligible studies were found through additional searches. This review contains results derived from 90 articles. The search results can be seen in Figure 1.
Figure 1. Study flow diagram.
Included studies
Details of included studies can be found in the Characteristics of included studies table in the full Cochrane publication. We included 90 studies containing 197,411 participants. Of these included studies, 88 were carried out in Africa, one in South America (Surinam), and one in Asia (Yemen).

Only one study was conducted in a hospital setting (antenatal clinic, outpatient setting). The other tests were performed in a field setting (village/school/military camp). *S. haematobium* was evaluated in most studies (n = 74); 16 evaluated *S. mansoni*. One study evaluated both species. Eighty studies reported the age of study participants; most of these were conducted in children (n = 50; 62.5%). Median prevalence of *S. haematobium* infection was 41% (range 1% to 89%), and that of *S. mansoni* infection was 36% (range 8% to 95%). Median female participation was 50% (Q1 46; Q3 53) for studies that reported gender (n = 46; 51%). Most of the included studies (n = 73; 81%) did not report on the status of praziquantel treatment in the study setting before the baseline study was performed.

Eighty-one studies used a cross-sectional design; six were cohort studies (longitudinal studies with follow-up), and three were case-control studies with controls from the same population (nested case-control studies). We included 84 English studies and six French studies. One study (Colley 2013), which was retrieved through an updated search, provided recent data for studies retrieved previously (Coulibaly 2011; Shane 2011; Tchuente 2012). In this case, we gathered data for the 2 × 2 tables from the most recent publication (Colley 2013).

Excluded studies
Full details of excluded studies can be found in the Characteristics of excluded studies table in the full Cochrane publication. We excluded 62 articles after reading the full texts. We excluded 17 case-control studies with healthy controls or with controls from non-endemic areas of schistosomiasis. We could not extract data from 2 × 2 tables for 16 studies. Twelve studies were not test accuracy studies, and four studies enrolled only patients proven to have schistosomiasis. Six studies used reference standards other than microscopy, four studies used other index tests to diagnose schistosomiasis that did not fulfill our inclusion criteria, and three studies performed similar tests on the same population as those reported by other already included studies.
Methodological quality of included studies (Figure 2) show results of the quality appraisal of the 60 included studies. Using the QUADAS-2 tool, we evaluated these studies for risk of bias in the following domains: participant selection, index test, reference standard, and participant flow. In general, poor reporting of quality items hindered our evaluation of quality. We therefore rated the risk of bias for these domains largely as unclear. In the participant selection domain, about 75% of studies were rated as having unclear risk of bias. For index tests, unclear risk of bias ranged from 80% to about 98% (about 98% for reagent strips for microhaematuria, about 95% for reagent strips for proteinuria, and about 80% for CCA POC testing). None of the studies had high risk of bias in the index test domain. For the reference standard, about 50% of the studies had high risk of bias, whereas the other half had unclear risk of bias. For the participant flow domain, about 75% of the studies had low risk of bias, and the remaining studies had unclear risk. Concerns for applicability for all four domains were predominantly low.

Findings
A summary of the main findings can be found in Summary of findings 1 and Summary of findings 2. Below we present in detail the overall findings for each index test.
Urine reagent strips

For microhaematuria
A total of 74 evaluations of the reagent strip for microhaematuria were performed with a total of 102,447 individuals. All evaluations were conducted in Africa. Median prevalence of S. haematobium was 42% (range 1% to 87%), and median female participation was 49% (Q1 49; Q3 53). Most of these evaluations were conducted with a lower-quality reference standard of only one slide/person (n = 63; 85%), and most evaluations were carried out in mixed populations of adults and children (n = 40; 54%). These evaluations were described in articles published between the years 1979 and 2014; a large proportion (n = 43; 58%) were published between 1979 and 1999. Over these four decades, no clear pattern was evident for effects of year of study on sensitivity and specificity of microhaematuria.

A large range of test brands were used to estimate the sensitivity and specificity of microhaematuria. Most evaluations (n = 25; 34%) were performed with the brand from the manufacturer Ames.

The forest plot (Figure 3) and the HSROC curve (Figure 4) for the reagent strip for microhaematuria reveal heterogeneity for estimates of both sensitivity and specificity. Meta-analytical sensitivity and specificity (95% confidence interval (CI)) of data at mixed thresholds were 75%(71% to 79%) and 87% (84% to 90%).
Figure 3. Forest plot of sensitivity and specificity of the urine reagent strip for microhaematuria. Squares represent sensitivity and specificity of one study, the black line its confidence interval.
For proteinuria

A total of 46 evaluations of the reagent strip for proteinuria were performed with a total of 82,113 individuals. All evaluations were conducted in Africa. Median prevalence of S. haematobium was 51% (range 4% to 89%), and median female participation was 50% (Q1 46; Q3 53). Most of these evaluations were conducted with a lower-quality reference standard (n = 36; 78%), and most were carried out in mixed populations of adults and children (n =28; 61%). These evaluations were described in articles published between the years 1979 and 2014; the largest proportion (n =27; 59%) were published before the year 2000. Over these four decades, no clear pattern was evident for effects of year of study on sensitivity and specificity of proteinuria.

A large range of test brands were used to estimate the sensitivity and specificity of proteinuria. Most evaluations (n = 17; 37%) were performed using the brand from the manufacturer Ames.
The forest plot (Figure 5) and the HSROC plot (Figure 6) for the reagent strip for proteinuria reveal greater heterogeneity for estimates of sensitivity than specificity. Meta-analytical sensitivity and specificity (95% CI) of data at mixed thresholds were 61% (53% to 68%) and 82% (77% to 88%).

Figure 5. Forest plot of sensitivity and specificity of the urine reagent strip for proteinuria. Squares represent the sensitivity and specificity of one study, the black line its confidence interval.
For leukocyturia

A total of five evaluations of the reagent strip for leukocyturia were performed with data from four publications and a total of 1532 individuals. Of these evaluations, two were carried out with a higher-quality reference standard (40%). Median prevalence of S. haematobium was 34% (range 4% to 77%), and median female participation was 100% (Q1 68; Q3 100). All evaluations except one were conducted in Africa in mixed populations of adults and children. These evaluations were described in articles published between the years 1992 and 2000; most (n = 3) were published before the year 2000. Two different test brands were evaluated. Most evaluations (n = 3; 60%) were done using the Nephur-test from Boehringer Mannheim. The forest plot (Figure 7) and the HSROC plot (Figure 8) for the reagent strip for leukocyturia reveal greater heterogeneity for estimates of specificity than sensitivity. The ROC plot also reveals poor accuracy of the test, as most study points lie close to the diagonal line. Meta-analytical sensitivity and specificity (95% CI) of data at mixed thresholds were 58% (44% to 71%) and 61%.
Figure 7. Forest plot of sensitivity and specificity of the urine reagent strip for leukocyturia.

Figure 8. Summary ROC plot of sensitivity versus specificity of the urine reagent strip for leukocyturia. The size of the points is proportional to the study sample size. The solid line shows the summary ROC curve.
Urine CCA POC test

For *S. haematobium*

A total of four evaluations of the CCA POC test for *S. haematobium* were performed on data derived from four publications with a total population of 901 individuals. Median prevalence of *S. haematobium* was 40% (range 31% to 48%), and median female participation was 47% (Q1 40; Q3 51). Most of these evaluations were conducted with a lower-quality reference standard (n = 3; 75%). All evaluations were conducted in Africa. All evaluations included data from children only. These evaluations were described in articles published between the years 2008 and 2011. Four different test brands were evaluated.

Forest plots (Figure 9) and ROC plots (Figure 10) for this test reveal a high degree of heterogeneity for estimates of both sensitivity and specificity. The ROC plot also reveals poor accuracy of the test, as the study points lie close to the diagonal line. Meta-analytical sensitivity and specificity (95% CI) of data at mixed thresholds were 39% (6% to 73%) and 78% (55% to 100%).

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Prevalence</th>
<th>RefStd</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bothard 2009a</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>102</td>
<td>Moderate</td>
<td>No</td>
<td>0.09 [0.02, 0.21]</td>
<td>0.96 [0.93, 1.00]</td>
</tr>
<tr>
<td>Ayeli 2008</td>
<td>51</td>
<td>39</td>
<td>47</td>
<td>69</td>
<td>Moderate</td>
<td>No</td>
<td>0.53 [0.42, 0.62]</td>
<td>0.44 [0.34, 0.57]</td>
</tr>
<tr>
<td>Metz 2009</td>
<td>64</td>
<td>68</td>
<td>23</td>
<td>70</td>
<td>Moderate</td>
<td>Yes</td>
<td>0.79 [0.70, 0.89]</td>
<td>0.44 [0.36, 0.52]</td>
</tr>
<tr>
<td>Ashton 2011</td>
<td>29</td>
<td>43</td>
<td>40</td>
<td>159</td>
<td>Moderate</td>
<td>No</td>
<td>0.37 [0.26, 0.49]</td>
<td>0.79 [0.72, 0.84]</td>
</tr>
</tbody>
</table>

*Figure 9. Forest plot of the sensitivity and specificity of the urine CCA POC test for *S. haematobium*. Squares represent the sensitivity and specificity of one study, the black line its confidence interval.*
For *S. mansoni*

A total of 15 evaluations of the CCA POC test for *S. mansoni* were performed on data derived from 13 publications with a total population of 6091 individuals. Median prevalence of *S. mansoni* was 36% (range 8% to 68%), and median female participation was 49% (Q1 48; Q3 51). Most of these evaluations were conducted with a lower-quality reference standard (n = 10; 67%). All evaluations were conducted in Africa, and all except one included data from children only. These 15 evaluations were described in articles published between the years 2007 and 2014. Two different test brands were evaluated: Rapid Diagnostic Tests from Pretoria South Africa and Schistosomiasis One Step Test from EVL Holland. Most evaluations (n = 9) were performed using the Rapid Diagnostic Tests from South Africa. The forest plot for this test reveals greater heterogeneity for estimates of specificity versus estimates of sensitivity (Figure 11). Meta-analytical sensitivity and specificity (95% CI) of data at a threshold ≥ trace positive were 89% (86% to 92%) and 55% (46% to 65%) (Figure 12).
Interpreting the Evidence-base for Making Recommendations on Medical Tests

Figure 11. Forest plot of sensitivity and specificity of the urine CCA POC test for S. mansoni. Squares represent the sensitivity and specificity of one study, the black line its confidence interval. Colley 2013 was a study that included data for 5 studies (done in different countries). Some of the studies had been published earlier (Coulibaly 2011, Erko 2013, Shane 2011, Tchuente 2012). In this case, we used data from Colley 2013, which provided the most recent and updated data.

Figure 12. Summary ROC plot of sensitivity versus specificity of the urine CCA POC test for S. mansoni. The size of the points is proportional to the study sample size. The thick black point shows the average value for sensitivity and specificity. The inner ellipse around the black spot represents the 95% confidence regions around the summary estimates. The outer ellipse represents the prediction region.
**For mixed infection**

One study assessed the capability of the POC test to detect schistosomiasis in an area of mixed *S. haematobium* and *S. mansoni* infection. This evaluation was conducted in Africa (Southern Sudan) in children only and was published in 2011. The brand used was Rapid Diagnostic Tests from Pretoria, South Africa. The sensitivity of the test was 66%, and the specificity was 79%. No meta-analysis was performed for this test because of insufficient data.

**CAA ELISA test**

**Serum**

A total of five evaluations of the serum CAA test for *S. mansoni* were performed on data derived from four publications (total population 1583, years of publication 1995 to 1998). Median prevalence of *S. mansoni* was 93% (range 28% to 96%), and median female participation was 49% (Q1 49; Q3 51). All of these evaluations were conducted using relatively higher-quality reference standards (n = 5; 100%). All were in-house assays, and one study involved only children. Sensitivity of the serum CAA ELISA for *S. mansoni* ranged from 47% to 94%, and specificity ranged from 8% to 100%. The ROC plot reveals a lot of scatter of the estimates of sensitivity and specificity provided by the included studies.

A total of three evaluations of the serum CAA test for *S. haematobium* were performed on data derived from three publications (total population 990, years of publication 1995 to 1999). Median prevalence of *S. haematobium* was 38% (range 18% to 57%). Only one study provided data on gender proportions (female participation was 54%). Two of the three evaluations were conducted using a higher-quality reference standard (67%). All were in-house assays, and all were carried out in mixed populations of adults and children. Sensitivity of the serum CAA test for *S. haematobium* ranged from 55% to 97%, and specificity ranged from 24% to 57%.

**Urine**

Only one evaluation of the urine CAA test for *S. mansoni* was performed on data derived from one publication (total population 204, year of publication 1995). This was an in-house assay and was done on data obtained from a mixed population of adults and children. Sensitivity of this test was 10%, and specificity was 99%. 
Only one evaluation of the urine CAA test for *S. haematobium* was performed on data derived from one publication (total population 370, year of publication 1999). This in-house assay was performed on data obtained from a mixed population of adults and children. Sensitivity of this test was 16%, and specificity was 94%.

**CCA ELISA test**

**Serum**

Two evaluations of the urine CCA test for *S. mansoni* were performed on data derived from two publications (total population 569, year of publication 1995). Both were in-house assays performed on data obtained from a mixed population of adults and children. Sensitivity of this test ranged from 36% to 85%, and specificity was 50% to 93%. Only one evaluation of the urine CCA test for *S. haematobium* was performed on data derived from one publication (total population 370, year of publication 1999). This in-house assay was performed on data obtained from a mixed population of adults and children. Sensitivity of this test was 3%, and specificity was 90%.

**Urine**

Two evaluations of the urine CCA test for *S. mansoni* were performed on data derived from two publications (total population 560, year of publication 1995). Both were in-house assays, and neither involved children only. Sensitivity of this test ranged from 62% to 97%, and specificity from 27% to 84%. Only one evaluation of the urine CCA test for *S. haematobium* was performed on data derived from one publication (total population 370, year of publication 1999). This in-house assay did not involve children only. Sensitivity of this test was 78%, and specificity was 70%.

**Comparisons of accuracy between reagent strips for microhaematuria and proteinuria**

Results of comparisons between microhaematuria and proteinuria are outlined in the Summary of findings. We first compared accuracy in all studies (indirect comparisons); we then limited the comparison to paired studies (direct comparisons). No statistically significant difference between the accuracy of microhaematuria and that of proteinuria was observed when the tests were compared in different populations using all studies (*P* = 0.25) (Figure 13). This can be demonstrated in the ROC curve showing the curves of tests as close together and crossing. The difference in accuracy also was not statistically significant when the tests were directly compared in the same individuals (*P* = 0.21) (Figure 14).
A statistically significant difference in the threshold parameter was noted when the tests were compared in different populations using all studies (P < 0.0001), and when the tests were directly compared in the same individuals (P = 0.0009). This could imply that one test has a different operating threshold when compared with the other, and although overall accuracy is not statistically significantly different, sensitivity and specificity may be different under field circumstances.

Figure 13. Summary ROC plot of sensitivity versus specificity showing the indirect comparison between microhaematuria and proteinuria (all studies). The solid lines show the summary ROC curves.
Figure 14. Summary ROC plot of sensitivity and specificity showing the direct comparison between microhaematuria and proteinuria (paired studies). Study points of microhaematuria and proteinuria.
Investigations of heterogeneity

Co-variates in the models
The co-variates quality of reference standard, age, gender (% female participation), prevalence of infection, and intensity of infection were added to the HSROC model. We investigated whether these co-variates affect the parameters of the HSROC model, that is, accuracy, threshold, and shape.

For the reagent strip for microhaematuria, the co-variates age ($P = 0.002$) and gender (% female participation) ($P = 0.02$) had statistically significant effects only on the threshold parameter of the HSROC model.

For the reagent strip for proteinuria, the co-variates quality of reference standard ($P = 0.01$) and prevalence of infection ($P$ value $0.007$) had statistically significant effects on the accuracy parameter. Accuracy was higher with the higher-quality reference standard and in settings with higher prevalence. Other co-variates did not have a statistically significant effect on any of the other parameters of the HSROC model.

For CCA POC used to detect S. mansoni, no co-variate had a statistically significant effect on sensitivity or on specificity.

Subgroup analysis
Table 1, Table 2, and Table 3 outline the results of subgroup analyses on the tests microhaematuria, proteinuria, and CCA POC for S. mansoni. When these tests were evaluated against the higher quality reference standard (ie when multiple samples were analyzed), sensitivity was lower for microhaematuria (71% vs 76%) and proteinuria (49% vs 68%) than with a lower-quality reference standard. Specificity of these tests was lower for microhaematuria (85% vs 87%) but higher for proteinuria (83% vs 78%). In contrast, sensitivity was similar (88%) and specificity was higher for the CCA POC test for S. mansoni (66% vs 55%) when measured against a higher-quality reference standard in comparison with a lower-quality reference standard.
### Table 1 Sources of heterogeneity for urine reagent strip for microhaematuria

<table>
<thead>
<tr>
<th>Group</th>
<th>Co-variates</th>
<th>Subgroup</th>
<th>n (N = 74)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall</strong></td>
<td></td>
<td></td>
<td></td>
<td>0.75 (0.71-0.79)</td>
<td>0.87 (0.84-0.90)</td>
</tr>
<tr>
<td><strong>Subgroup analysis</strong></td>
<td>Reference standard</td>
<td>Higher quality (&gt; 1 sample)</td>
<td>10</td>
<td>0.71 (0.62-0.80)</td>
<td>0.85 (0.78-0.93)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lower quality (1 sample)</td>
<td>64</td>
<td>0.76 (0.71-0.80)</td>
<td>0.87 (0.84-0.90)</td>
</tr>
<tr>
<td></td>
<td>Threshold ≥ &gt;1</td>
<td></td>
<td>23</td>
<td>0.80 (0.73-0.85)</td>
<td>0.85 (0.78-0.92)</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>Children</td>
<td>34</td>
<td>0.77 (0.71-0.82)</td>
<td>0.91 (0.87-0.93)</td>
</tr>
<tr>
<td></td>
<td>Intensity of infection</td>
<td>Light</td>
<td>28</td>
<td>0.73 (0.66-0.79)</td>
<td>0.88 (0.84-0.92)</td>
</tr>
<tr>
<td><strong>Sensitivity analysis</strong></td>
<td>Concentration</td>
<td>Filtration only</td>
<td>62</td>
<td>0.73 (0.69-0.78)</td>
<td>0.86 (0.82-0.89)</td>
</tr>
<tr>
<td></td>
<td>QUADAS Patient Selection</td>
<td>Low risk of bias</td>
<td>16</td>
<td>0.77 (0.70-0.86)</td>
<td>0.86 (0.79-0.92)</td>
</tr>
<tr>
<td></td>
<td>QUADAS Reference Standard</td>
<td>Low risk of bias*</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>QUADAS Flow and Timing</td>
<td>Low risk of bias</td>
<td>43</td>
<td>0.77 (0.72-0.82)</td>
<td>0.87 (0.83-0.90)</td>
</tr>
</tbody>
</table>

*Insufficient data for synthesis.
### Table 2 Sources of heterogeneity for urine reagent strip proteinuria

<table>
<thead>
<tr>
<th>Group</th>
<th>Co-variates</th>
<th>Subgroup</th>
<th>n (N=46)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td>0.61 (0.53-0.68)</td>
<td>0.82 (0.77-0.88)</td>
</tr>
<tr>
<td>Subgroup analysis</td>
<td>Reference standard</td>
<td>Higher quality (&gt;1 sample)</td>
<td>9</td>
<td>0.49 (0.28-0.70)</td>
<td>0.83 (0.76-0.90)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lower quality (1 sample)</td>
<td>37</td>
<td>0.68 (0.60-0.76)</td>
<td>0.78 (0.69-0.87)</td>
</tr>
<tr>
<td></td>
<td>Threshold</td>
<td>≥ +1</td>
<td>13</td>
<td>0.69 (0.56-0.81)</td>
<td>0.72 (0.54-0.90)</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>Children</td>
<td>18</td>
<td>0.67 (0.56-0.76)</td>
<td>0.81 (0.74-0.87)</td>
</tr>
<tr>
<td></td>
<td>Intensity of infection</td>
<td>Light</td>
<td>15</td>
<td>0.60 (0.43-0.77)</td>
<td>0.83 (0.73-0.93)</td>
</tr>
<tr>
<td>Sensitivity analysis</td>
<td>Concentration</td>
<td>Filtration only</td>
<td>35</td>
<td>0.62 (0.52-0.71)</td>
<td>0.80 (0.73-0.86)</td>
</tr>
<tr>
<td></td>
<td>QUADAS</td>
<td>Patient Selection</td>
<td>Low risk of bias</td>
<td>11</td>
<td>0.64 (0.50-0.79)</td>
</tr>
<tr>
<td></td>
<td>QUADAS</td>
<td>Reference Standard</td>
<td>Low risk of bias</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>QUADAS Flow and Timing</td>
<td>Low risk of bias</td>
<td>36</td>
<td>0.67 (0.59-0.76)</td>
<td>0.82 (0.73-0.88)</td>
</tr>
</tbody>
</table>

*Insufficient data for synthesis.

### Table 3. Sources of heterogeneity for CCA POC test for S. mansoni

<table>
<thead>
<tr>
<th>Group</th>
<th>Co-variates</th>
<th>Subgroup</th>
<th>n (N=15)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td>0.89 (0.86-0.92)</td>
<td>0.55 (0.46-0.65)</td>
</tr>
<tr>
<td>Subgroup analysis</td>
<td>Reference standard</td>
<td>Higher quality (&gt;1 sample)</td>
<td>5</td>
<td>0.88 (0.82-0.92)</td>
<td>0.66 (0.46-0.82)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lower quality (1 sample)</td>
<td>13</td>
<td>0.88 (0.85-0.91)</td>
<td>0.55 (0.45-0.66)</td>
</tr>
</tbody>
</table>
Microhaematuria and proteinuria had higher sensitivity (77% vs 73% and 67% vs 56%) in children than in mixed populations of adults and children. Specificity was higher for microhaematuria (91% vs 82%) but was comparable for proteinuria (81% vs 82%) in children compared with mixed populations of adults and children. All except one study of CCA POC for S. mansoni were carried out with children. At a positivity threshold ≥ 1, sensitivity of CCA POC for S. mansoni was lower (72% vs 89%) and specificity higher (85% vs 55%) than at a positivity threshold of trace positive. In the light-intensity subgroup, sensitivity was slightly lower for microhaematuria (73% vs 75%) and specificity was slightly higher (88% vs 87%) compared with results of the overall analysis. In contrast, sensitivity (60% vs 61%) and specificity (83% vs 82%) for proteinuria were comparable. Data were insufficient to permit estimation of the sensitivity and specificity of CCA POC for S. mansoni in light-intensity settings.

The forest plot (Figure 15) and the ROC plot (Figure 16) demonstrating sensitivity and specificity for microhaematuria after praziquantel treatment show a lot of variation in the estimates (predominantly for sensitivity) of the individual studies.
Figure 15. Forest plot of sensitivity and specificity of the urine reagent strip for microhaematuria for studies done after treatment with praziquantel. Squares represent the sensitivity and specificity of one study, the black line its confidence interval.

![Forest Plot]

Figure 16. Summary ROC plot of sensitivity and specificity of the urine reagent strip for microhaematuria for studies done after treatment with praziquantel. The size of the points is proportional to the study sample size.

![Summary ROC Plot]
Sensitivity analysis
For microhaematuria, when the analysis was limited to studies that used filtration only as the concentration method for urine microscopy, sensitivity (73% (69% to 78%) vs 76% (72% to 80%)) was lower and specificity was comparable (86% (82% to 89%) vs 86% (82% to 89%)) with those produced by the overall analysis. For proteinuria, when the analysis was limited to studies that used filtration only as the concentration method for urine microscopy, sensitivity was comparable (62% (52% to 71%) vs 61% (53% to 69%)) and specificity was lower (80% (73% to 86%)) than those produced by the overall analysis (83% (77% to 88%)) (Table 1; Table 2; Table 3).

Sensitivities and specificities of microhaematuria were comparable when analysis was limited to studies with low risk of bias for the participant flow domain. Sensitivity of proteinuria was higher when limited to studies with low risk of bias for the participant selection domain (64%) and the participant flow domain (67%). Specificity on the other hand was comparable for these two domains. Sensitivity and specificity of CCA POC for S. mansoni were comparable when limited to studies with low risk of bias for the participant flow domain (Table 1; Table 2; Table 3). Data were insufficient to allow estimation of sensitivity and specificity for studies with low risk of bias in the other domains-reference standard and participant selection-for the CCA POC test for S. mansoni.

As part of post hoc analyses, we noted that three evaluations showed substantial heterogeneity for the tests microhaematuria (Aryeetey 2000; sensitivity 55%, specificity 36%), proteinuria (Aryeetey 2000; sensitivity 38%, specificity 11%), and CCA POC for S. mansoni (Standley 2010; sensitivity 99%, specificity 19%). We excluded these evaluations in sensitivity analyses for the respective tests and found the following results. Results for microhaematuria (sensitivity 75%, specificity 87%) and proteinuria (sensitivity 61%, specificity 82%) were similar to those of the overall analysis. For CCA POC for S. mansoni, sensitivity was comparable (88% vs 89%) and specificity was slightly higher (58% vs 55%) compared with those of the overall analysis.
<table>
<thead>
<tr>
<th>What is the diagnostic accuracy of circulating antigen tests and biochemical urine reagent strips in detecting <em>S. haematobium</em> infection?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patients/Population</strong></td>
</tr>
<tr>
<td><strong>Prior treatment with praziquantel before baseline study</strong></td>
</tr>
<tr>
<td><strong>Prior testing</strong></td>
</tr>
<tr>
<td><strong>Settings</strong></td>
</tr>
<tr>
<td><strong>Index tests</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Reference standard</strong></td>
</tr>
<tr>
<td><strong>Importance</strong></td>
</tr>
<tr>
<td><strong>Studies</strong></td>
</tr>
<tr>
<td><strong>Quality concerns</strong></td>
</tr>
<tr>
<td><strong>Test types</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>---</td>
</tr>
</tbody>
</table>
### Interpreting the Evidence-base for Making Recommendations on Medical Tests

#### Summary of findings 1.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Count 1</th>
<th>Count 2</th>
<th>Count 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biochemical urine reagent strips</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For microhaematuria</td>
<td>75% (71% to 91%)</td>
<td>87% (84% to 90%)</td>
<td>102</td>
<td>77</td>
<td>384</td>
</tr>
<tr>
<td>For proteinuria</td>
<td>61% (53% to 68%)</td>
<td>82% (77% to 88%)</td>
<td>160</td>
<td>106</td>
<td>356</td>
</tr>
<tr>
<td>For leukocyturia</td>
<td>58% (44% to 71%)</td>
<td>61% (34% to 88%)</td>
<td>172</td>
<td>230</td>
<td>468</td>
</tr>
<tr>
<td><strong>Circulating cathodic antigen test (CCA)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urine POC test</td>
<td>39% (6% to 73%)</td>
<td>78% (55% to 100%)</td>
<td>250</td>
<td>94</td>
<td>254</td>
</tr>
</tbody>
</table>

#### Comparisons

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Comparison type</th>
<th>Number of evaluations and differences in overall accuracy</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microhaematuria vs proteinuria</td>
<td>All studies</td>
<td>74 microhaematuria proteinuria, difference in accuracy ($P = 0.25$)</td>
<td>Proteinuria would be expected to miss 14% more cases than microhaematuria</td>
</tr>
<tr>
<td></td>
<td></td>
<td>We found no evidence of a statistically significant difference in overall accuracy when microhaematuria and proteinuria are carried out and compared</td>
<td>Proteinuria would be expected to falsely identify 5% more cases than microhaematuria</td>
</tr>
</tbody>
</table>
Summary of findings 1.

| Paired studies (tests done in the same individuals) | 44 microhaematuria vs proteinuria, differences in accuracy (P = 0.21) | We found no evidence of a statistically significant difference in overall accuracy when microhaematuria and proteinuria are carried out and compared in the same individuals |

* Studies were insufficient to provide summary estimates for the CAA tests.

When the tests were evaluated against the higher-quality reference standard (ie when multiple samples were analyzed), sensitivity was lower for microhaematuria (71% vs 76%) and proteinuria (49% vs 61%) in comparison with a lower-quality reference standard. The specificity of these tests was comparable.

In light-intensity settings, sensitivity was slightly lower for microhaematuria (73% vs 76%) and specificity was slightly higher (88% vs 86%) compared with results of the overall analysis. In contrast, sensitivity (60% vs 61%) and specificity (83% vs 83%) for proteinuria were comparable.

Microhaematuria and proteinuria had higher sensitivity (77% vs 73% and 67% vs 56%) in children than in mixed populations of adults and children. Specificity was higher for microhaematuria (91% vs 82%) but specificity was comparable for proteinuria (81% vs 82%) in children compared with mixed populations of adults and children.

For the effects of risk of bias, sensitivities and specificities of microhaematuria were comparable when limited to studies with low risk of bias for the participant flow domain. Sensitivity of proteinuria was higher when limited to studies with low risk of bias for the participant selection domain (64%) and the participant flow domain (67%). Specificity on the other hand was comparable for these 2 domains.

Abbreviations: TPs (true-positives), FPs (false-positives), FNs (false-negatives).
What is the diagnostic accuracy of circulating antigen tests for *S. mansoni* infection?

<table>
<thead>
<tr>
<th>Patients/Population</th>
<th>People residing in areas endemic for <em>S. mansoni</em> infection (16 out of 90 studies)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior treatment with praziquantel before baseline study</td>
<td>Yes (1 study), No (5 studies), Unclear (10 studies)</td>
</tr>
<tr>
<td>Prior testing</td>
<td>None</td>
</tr>
<tr>
<td>Settings</td>
<td>Field settings (villages, schools, and military camp) in Africa and South America</td>
</tr>
<tr>
<td>Index tests</td>
<td>Circulating cathodic antigen test (CCA)</td>
</tr>
<tr>
<td></td>
<td>Circulating anodic antigen test (CAA)</td>
</tr>
<tr>
<td>Reference standard</td>
<td>Stool microscopy</td>
</tr>
</tbody>
</table>

**Importance**

These tests are being used as replacements for conventional microscopy in disease control programmes for schistosomiasis, as they are rapid, are easier to use and interpret, and may have comparable sensitivity to microscopy. As control programmes gain impetus and infection intensities decrease, higher sensitivities become a prerequisite for future diagnostics.

**Studies**

Cross-sectional studies

**Quality concerns**

Poor reporting of participant characteristics, index test and reference standard methods, and intensity of infection were common concerns. The risk of bias assessment for most included studies was largely unclear for the QUADAS domains Patient Selection, Index Tests, and Reference Tests.

<table>
<thead>
<tr>
<th>Test types</th>
<th>Number of evaluations</th>
<th>Summary estimates (95% CI)</th>
<th>In 1000 people tested</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Circulating cathodic antigen test (CCA)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary of findings 2.

| Urine POC test | 15 | Sens = 89% (86% to 92%), Spec = 55% (46% to 65%) | 40 | 283 | 608 |

* Studies were insufficient to provide summary estimates for CAA tests.
  When measured against a higher-quality reference standard, sensitivity of CCA POC for *S. mansoni* was comparable (88% vs 88%) but specificity was higher (66% vs 55%) than when measured against a lower-quality reference standard.
  At a positivity threshold ≥ 1, sensitivity of CCA POC for *S. mansoni* was lower (72% vs 87%) and specificity higher (85% vs 61%) than at a positivity threshold of trace-positive.
  Data were insufficient to estimate the sensitivity of CCA POC for *S. mansoni* in light-intensity settings.
  For the effects of risk of bias, sensitivity and specificity of CCA POC for *S. mansoni* were comparable when limited to studies with low risk of bias for the participant flow domain.

Abbreviations: TPs (true-positives), FPs (false-positives), FNs (false-negatives).
Discussion

Summary of main results
This review focused on analyzing the accuracy of urine reagent strips for the diagnosis of S. haematobium and of circulating antigen tests for the detection of S. haematobium and S. mansoni infections. Microscopy was used as the reference standard, and 90 studies were found to fit our inclusion criteria; data from these studies were used in this review. The main results, including average sensitivities and specificities for tests included in the metaanalyses, are reported in Summary of findings 1 and Summary of findings 2.

Most of the studies included in our overall meta-analyses used a 'lower-quality reference test': microhaematuria 81%, proteinuria 73%, leukocyturia 60%, circulating cathodic antigen point-of-care (CCA POC) for S. haematobium 75%, and CCA POC for S. mansoni 81%. This implies that infections missed by single ample microscopy may have increased the number of false-positives identified by the index tests, consequently leading to lower estimates of specificity.

Our overall analyses suggest that among the tests used to detect S. haematobium, the urine reagent strip for microhaematuria detects the largest proportion of schistosome infections identified by microscopy (sensitivity 75%); it also detects the largest proportion of non-infections identified by microscopy (specificity 87%). Proteinuria follows suit, with sensitivity of 61% and specificity of 82%.

The superior performance of microhaematuria over proteinuria was not statistically significant when the comparison was performed both indirectly (using all studies) and directly (using paired studies) within the HSROC model. When measured against a higher-quality reference standard (multiple measurements), microhaematuria had both lower sensitivity (71% vs 75%) and lower specificity (85% vs 87%) than were seen with a lower-quality reference standard. Proteinuria on the other hand, when measured against a higher-quality reference standard, had lower sensitivity (49% vs 61%) and higher specificity (82% vs 78%) versus a lower quality reference standard. Increasing the sensitivity of microscopy by taking multiple measurements may reduce the number of true cases wrongly classified as non-infected by microscopy.

An index test compared against a more sensitive reference test (higher quality) may have higher specificity because the number of false-positives will be low. The lower specificity for microhaematuria may be due in part to poor reporting of how the reference standard way conducted in some studies.
Our results suggest that the urine reagent strip when used to detect leukocyturia is limited by low sensitivity (58%) and specificity (61%) and is not useful in practice. The low sensitivity for leukocyturia could be explained by the variations in morbidity caused by S. haematobium. Not all infected people have leukocyturia; therefore the proportion of false-negatives is higher. The CCA POC test has very low sensitivity (39%) to detect S. haematobium and specificity of 78% and may not be suitable for mapping or estimation of infection, because it misses very many infections identified by microscopy.

The CCA POC test for S. mansoni detected a large proportion of infections identified by microscopy (sensitivity 89%). However, it also detected a lower proportion of the non-infected cases identified by microscopy (specificity 55%). The low specificity can be explained by the fact that most studies in the overall analyses were measured against a lower-quality reference standard. When compared with a higher-quality reference standard, the CCA POC test had comparable sensitivity (88%) but higher specificity (66%). Arguably, if the reference standard had been even better, this specificity might have increased further.

As studies were insufficient, we were unable to generate summary estimates for the circulating antigen enzyme-linked immunosorbent assay (ELISA) tests (CCA and circulating anodic antigen (CAA)). Estimates of sensitivity and specificity from the included studies evaluating these tests ranged widely. Results of our assessment of risk of bias of the included studies were largely unclear because of poor reporting of items in these studies.

**Application of the meta-analysis to a hypothetical cohort**

Summary of findings 1 and Summary of findings 2 apply the results of the meta-analyses to a hypothetical cohort of 1000 individuals suspected of having active *S. haematobium* and/or active *S. mansoni* infection in a field setting. We illustrate the impact of using microhaematuria, proteinuria, leukocyturia, and CCA POC for *S. haematobium* in a setting with a prevalence of *S. haematobium* infection of 41%, and the impact of using CCA POC for *S. mansoni* in a setting with a prevalence of *S. mansoni* infection of 36%. These are the estimates of median prevalence of infection obtained from all studies included in this review.
Delivery of population-based control programmes such as treatment with praziquantel requires knowledge of prevalence estimates of schistosomal infections (Colley 2014). This helps the clinician in determining whether mass drug treatment should be administered in settings of very high prevalence, or targeted treatment in settings of low prevalence. We have included descriptions of the performance of these tests in estimating the prevalence (index test positives (TP + FP)) of *S. haematobium* and *S. mansoni* infections.

**S. haematobium** infection

If the point estimates of the tests for *S. haematobium* are applied to a hypothetical cohort of 1000 individuals suspected of having active *S. haematobium* infection, among whom 410 actually have the infection, the strip for microhaematuria would be expected to miss (102) and falsely identify (77) the least number of cases. This test would identify 384 positive cases in total.

For the other tests (in increasing order of missed cases): The strip for proteinuria would be expected to miss 160 cases and to falsely identify 106 cases; proteinuria would be expected to miss 14% more cases than microhaematuria and to falsely identify 5% more cases than microhaematuria; leukocyturia would be expected to miss 172 cases and to falsely identify 230 cases; and the CCA POC test would be expected to miss 250 cases and to falsely identify 130 cases. In total, the strips for proteinuria, leukocyturia, and the CCA POC test would identify 356, 468, and 254 positive cases, respectively.

Overall, when infection is mapped, the prevalence of microhaematuria would seem to be 38%—close to the true prevalence of 41%. The prevalence of proteinuria would seem to be 36%, that of leukocyturia 47%, and that of CCA POC 25%. In cases of mass treatment, the ultimate consequences of these numbers would depend on the minimal prevalence needed to start mass treatment.

**S. mansoni** infection

If the point estimates for the CCA POC test are applied to the same hypothetical cohort of 1000 individuals suspected of having active *S. mansoni* infection among whom 360 actually have the infection, the CCA POC test would be expected to miss 40 cases and to falsely identify 288 cases. In total, the test would identify 608 positive cases (for an observed prevalence of 61%).
Comparison with other reports
The absence of a suitable gold standard for active schistosomiasis is reflected in the existing literature, where different reference standards are used with subsequent variation in accuracy (especially with specificity) of the index test (Koukounari 2009; Coulibaly 2011; Tchuente 2012; Colley 2013; Erko 2013; King 2013; Lodh 2013; Sousa-Figueiredo 2013).

A meta-analysis was recently published that assessed the accuracy of urine reagent strips for microhaematuria against conventional microscopy as a reference standard (King 2013). Unlike King’s review, our review also estimated the accuracy of other urine reagent strips for proteinuria and leukocyturia. To guide decision making, it is important to show which of these tests fares better. Our analyses suggest that microhaematuria has higher sensitivity than proteinuria and leukocyturia.

Compared with results from King’s meta-analysis (King 2013), our estimate of sensitivity for microhaematuria was lower (75% vs 81%) but specificity was comparable (87% vs 89%). This difference may be attributed to the method of meta-analysis used. King used the HSROC regression following a Bayesian Monte Carlo Markov chain approach (Dendukuri 2012), and we used the HSROC model recommended in the Cochrane Handbook for Systematic Reviews of Interventions (Macaskill 2010). With regard to sources of heterogeneity, some of our results are also comparable with those of King 2013. For instance, King found through multi-variable regression modelling that the urine heme dipstick performed better in children than in mixed populations of adults and children (Relative diagnostic odds ratio = 3.16). In our review, we found that sensitivity and specificity were higher in studies on children compared with studies on mixed populations of adults and children. We strongly confirm that this test is therefore highly suitable for mass mapping of school-aged children in endemic areas. Again our analyses show that sensitivity of the urine heme dipstick was slightly lower in settings of low intensity (73%) compared with that of the overall estimate (75%). This finding was similar to the findings of King, which showed that sensitivity of the urine heme dipstick was lower in settings of lower infection intensity (65%) in the subgroup analysis than in the overall analysis (81%). However it should be noted that our definition of light intensity differed from that used by King. We selected the more commonly used World Health Organization (WHO) recommended cutoff of < 50 eggs per 10 mL, whereas King defined low intensity as ≤ 100 eggs/10 mL. This could explain in part why our sensitivity estimates were higher than those of King in settings of light intensity.
A key difference between our review and that of King 2013 concerned the effects of treatment on the estimate of sensitivity of the heme dipstick. In a subgroup of eight studies with mixed posttreatment evaluations of one year (n = 6), six months (n = 1), and one month (n = 1), King’s review produced a lower summary estimate of sensitivity (72%) in the subgroup of treated populations as compared with the overall analysis (81%). King considered treatment evaluations with praziquantel and metrifonate, whereas we focused on studies that evaluated the effects of praziquantel treatment, as this is the current drug of choice. Because studies reported varied time intervals between treatment and retesting, we opted not to pool the estimates of studies, as this would likely produce biased overestimates of sensitivity and specificity.

Studies with long time intervals were likely to include greater numbers of participants reinfected compared with studies carried out at shorter time intervals, and their results may be confounded by repeated treatments provided by national programmes.

A recently published multi-centre evaluation of CCA POC tests done in five African countries (Colley 2013) recommended that the CCA POC test for S. mansoni (evaluated with a positivity threshold ≥ trace positive) was a sufficiently sensitive and specific tool for mapping intestinal schistosomiasis in moderate- to high-prevalence areas, and therefore it was a viable alternative to microscopy (Colley 2013).

After acknowledging the absence of a gold standard, this multi-centre study used latent class analysis (modelling results from CCA POC, Kato-Katz, and PCR) to generate an overall estimate of 86% sensitivity and 72% specificity of the CCA POC based on data from 4405 school-age children. Using microscopy only (KK) as the reference standard, our review, which incorporated all included study results along with findings of additional studies, produced a comparable summary estimate of 89% sensitivity but a lower summary estimate of 55% specificity at a threshold of trace positive. Differences in specificity could be explained by the reference standard and indicate that some of the false-positives identified by CCA POC are indeed likely to be true infections that are not detected by standard microscopy.

Few studies have fully evaluated the accuracy of the circulating antigen ELISA tests (CCA and CAA). The serum CAA ELISA test is currently being converted to a point-of-care format for S. mansoni (Corstjens 2008) and S. haematobium (van Dam 2013) with promising results of analytical sensitivity and specificity. In our review, sensitivity of the included studies evaluating the serum CAA ELISA test for S. mansoni ranged widely from 47% to
94%, and specificity ranged widely from 8% to 100%. Sensitivity of the included studies evaluating the serum CAA ELISA test for *S. haematobium* ranged from 55% to 97%, and specificity was low, ranging from 24% to 57%. However, the studies included in our review were carried out before the year 2000 with in-house tests. The tests currently being developed are most likely improved versions; therefore additional studies analyzing the clinical sensitivity and specificity of the serum ELISA tests are needed for conclusive determination of whether they are suitable for the diagnosis of active schistosomiasis.

**Strengths and weaknesses of the review**

**Strengths**

We have evaluated the accuracy of POC tests currently in use and tests that have recently been transformed into POC tests for detection of active schistosomiasis in endemic areas. This makes our review relevant to current practice. To avoid missing studies, we did not use a search filter, and we did not limit our search by publication year or language; also to limit bias, data extraction was performed by two people independently.

**Weaknesses**

**Choice of the reference standard**

In light of the absence of a suitable gold standard for active schistosomiasis and the presence of other proposed alternative reference standards, evaluation of index tests with only microscopy as the reference standard may be considered a shortcoming of our review. However because microscopy remains the most commonly used test and therefore reference test, we wanted our review to be applicable to current practice. Our review provides better insight into the proportion of cases detected and the proportion of cases misclassified by urine reagent strips and CCA POC tests when microscopy is used as the reference standard.

A more reliable way of evaluating whether an index test can replace microscopy would be to compare the accuracy of microscopy, urine reagent strips, and circulating antigen tests against other proposed reference standards in the same set of participants (direct comparison studies). A few studies have compared the accuracy of one or more KK smears and CCA POC against a reference standard comprising six or more KK smears (Coulibaly 2011; Tchuente 2012; Erko 2013) or against PCR as the reference test (Lodh 2013).

All of these studies have shown the CCA POC test to be more sensitive but less specific.
than single or double KK. More direct comparative studies and reviews are needed to reliably confirm this finding and to identify sources of variation in results.

**Quality of included studies**

Poor and inconsistent reporting of participant characteristics such as clinical status of participants, intensity of infection, administration of praziquantel treatment, and conduct of the study limited our investigations of sources of heterogeneity and risk of bias assessment.

In our review, the reporting of intensity of infection was unclear (reported in different ways (arithmetic mean or range of infection or geometric mean or range of infection or proportions with light/moderate/heavy infections) or not reported at all) for a large proportion of the included studies (microhaematuria 44%, proteinuria 42%, and CCA POC 45%). It was therefore difficult to effectively investigate its influence on the accuracy of the evaluated tests.

It was also a challenge to fully investigate the effects of praziquantel treatment on the accuracy of the evaluated tests because 82% of the studies did not report the treatment status of participants before the start of the study. The effects of intensity of infection and the effects of praziquantel treatment on the accuracy of diagnostic tests for schistosomiasis are currently an important concern for national control programmes, particularly as praziquantel treatments progress, with subsequent decreases in infection intensities. Indeed, in areas where the force of infection and associated morbidity have been greatly reduced, some programmes are beginning to focus on elimination. It is therefore of vital importance that highly sensitive tests are used for monitoring, and that highly sensitive and specific tests are used in efficacy studies before and after treatment.

**Applicability of findings to the review question**

Our concern about the applicability of the included studies to our review question was low, as assessed by QUADAS-2. As all but one study were carried out in Africa, and all but one study were conducted in field settings, our results are highly applicable for use in endemic communities for which disease control programmes are often targeted. However, one area that may limit the applicability of our findings to the review question is our investigation into sources of heterogeneity such as effects of praziquantel treatment and risk of bias assessment on the accuracy estimates of evaluated tests.

As discussed earlier, poor and inconsistent reporting limited this investigation. In light of the ongoing disease control programmes, fully showing any variation in test accuracy associated with effects of praziquantel treatment would be useful for policy makers.
Knowing the risk of bias of included studies would also help in objective assessments of the strength of the evidence. Study authors therefore are encouraged to use the Standards for Reporting of Diagnostic Accuracy Studies (STARD) guidelines (Bossuyt 2003) in reporting the design and conduct of their studies.

**Authors’ conclusion**

**Implications for practice**

Among the tests evaluated for *S. haematobium* infection, microhaematuria has detected the largest proportion of infections and non-infections identified by microscopy. This test could continue to serve as a replacement test for microscopy for initial mapping or estimation of *S. haematobium* infection, particularly in endemic areas with moderate to high prevalence of infection.

The CCA POC test for *S. mansoni* detects a very large proportion of infections identified by microscopy but misclassifies many microscopy negatives as positives in endemic areas with moderate to high prevalence of infection. This may occur because the test is potentially more sensitive than microscopy. Nevertheless, healthcare workers should interpret the results with care when using this test for initial mapping or estimation of *S. mansoni* infection, as some of the positives may still be false-positives, in particular when trace-positive is used as the threshold.

Besides assessment of the accuracy of a test, the choice of a suitable diagnostic test should be made in light of cost and logistical considerations. Costs for microscopy (USD per examination, 0.3 for a single thick KK smear) (Cavalcanti 2013) and for reagent strips for microhaematuria (USD 0.32) (Legesse 2008) are comparable, but the strips are easier to use and interpret and therefore are not logistically challenging in field settings. The CCA POC tests are more costly (USD 2.6 per examination) (Cavalcanti 2013) but are rapid and easy to use and interpret, are highly portable, and require fewer technical personnel than microscopy; they are also suitable for field screening and diagnosis.

**Implications for research**

As control programmes progress with expected subsequent decreases in prevalence and intensity of infection, we highlight the importance of additional primary research conducted to identify a suitable clinical reference standard for active schistosomiasis.
Additional studies comparing the accuracy of microscopy, circulating antigen tests, and urine reagent strips versus other proposed reference standards are needed if a suitable replacement for microscopy in practice is to be reliably recommended. Further studies to identify other sensitive tests to detect active *S. haematobium* and *S. mansoni* infections and further evaluations of the CAA test as a future POC test for serum or urine are also needed.

For suitable tests to be reliably recommended for monitoring effects of praziquantel treatment in disease control programmes, additional follow-up studies are required to evaluate the effects of praziquantel treatment on intensity of infection and accuracy of urine reagent strips and circulating antigen tests. Further research on cost-effectiveness of diagnostic tests in areas of different endemicity is also needed, as cost is a key deciding factor in resource-limited settings.

Finally, authors of primary test accuracy studies should be encouraged to use the STARD guidelines when reporting the design and conduct of their studies. This will enable systematic reviewers to better synthesize the data and to draw conclusions on risk of bias in studies of test accuracy.

**Acknowledgement**

We thank René Spijker, MSc (Dutch Cochrane Centre, University of Amsterdam), for assisting in the development of the search strategy of this project.
References to studies included in this review

Abdel-Wahab 1992 [published data only]

Abdel-Wahab 2000 [published data only]

Adriko 2014˙6KK [published data only]

Adriko 2014˙settingA [published data only]

Adriko 2014˙settingB [published data only]

Adriko 2014˙settingC [published data only]
Alsherbiny 1999 (published data only)

Anosike 2001 (published data only)

Aryeetey 2000 (published data only)

Ashton 2011 (published data only)

Ayele 2008 (published data only)

Bassiouny 2014 (published data only)

Birrie 1995’settingA (published data only)

Birrie 1995’settingB (published data only)
Birrie 1995 (published data only)

Bogoch 2012 (published data only)

Bosompem 1996 (published data only)

Bosompem 2004 (published data only)

Colley 2013 (Uganda) (published data only)

Cooppan 1987 (published data only)

Coulibaly 2011 (9KK) (published data only)

Coulibaly 2011 Colley 2013 (published data only)
Coulibaly 2013 (published data only)

De Clerq 1995 (published data only)

El-Morshedy 1996 (published data only)

El-Sayed 1995 (published data only)

Eltoum 1992 (published data only)

Erko 2013 (published data only)

Erko 2013 Colley 2013 (published data only)
Etard 2004 {published data only}  

Fatiregun 2005 {published data only}  

French 2007 {published data only}  

Gabr 2000 {published data only}  

Gigase 1988 {published data only}  

Gundersen 1996 {published data only}  

Hall 1999 {published data only}  

Hammam 2000’a {published data only}  
Interpreting the Evidence-base for Making Recommendations on Medical Tests

Hammam 2000b [published data only]

Houmsou 2011 [published data only]

Kassim 1989 [published data only]

Kiliku 1991 [published data only]

King 1988a [published data only]

King 1988b [published data only]

Kitange 1993 [published data only]

Legesse 2007 [published data only]
Legesse 2008 (published data only)

Lengeler 1993 (published data only)

Mafe 1997 (published data only)

Mafe 2000 (published data only)

Magnussen 2001 (published data only)

Midzi 2009 (published data only)

Morenikeji 2014 (published data only)

Mott 1985a’1 (published data only)
Mott 1985a \cite{Mott1985} (published data only)

Mtasiwa 1996 (published data only)

Murare 1987 (published data only)

Navaratnam 2012 (published data only)

Ndamukong 2001 (published data only)

Ndlovu 1996 (published data only)

Nduka 1995 (published data only)

Ndumugyenyi 2001 (published data only)
Ngându 1988 \(\text{[published data only]}\)

NGoran 1989 \(\text{[published data only]}\)

Nmorsi 2005 \(\text{[published data only]}\)

Nwaorgu 1992 \(\text{[published data only]}\)

Ofori 1986 \(\text{[published data only]}\)

Okeke 2014˙settingA \(\text{[published data only]}\)

Okeke 2014˙settingB \(\text{[published data only]}\)

Onayade 1996 \(\text{[published data only]}\)
Interpreting the Evidence-base for Making Recommendations on Medical Tests

Poggensee 2000˙settingA (published data only)

Poggensee 2000˙settingB (published data only)

Polman 1995 (published data only)

Pugh 1980 (published data only)

Rasendramino 1998 (published data only)

Robinson 2009 (published data only)

Rollinson 2005 (published data only)

Sarda 1985 (published data only)
Sarda 1986 (published data only)

Savioli 1990 (published data only)

Sellin 1982 (published data only)

Shane2011˙Colley2013 (published data only)

Shaw 1998 (published data only)

Standley 2010 (published data only)

Stephenson 1984 (published data only)

Stothard 2006 (published data only)
Stothard 2009a (published data only)

Stothard 2009b (published data only)

Tanner 1983’1 (published data only)

Tanner 1983’2 (published data only)

Tchuente 2012’9KK (published data only)

Tchuente 2012’Colley2013 (published data only)

Traore 1998 (published data only)

Ugomoiko 2009a (published data only)

Ugomoiko 2009b’1 (published data only)
Ugbomoiko 2009b {published data only}

Van Lieshout 1995 {published data only}

Van Lieshout 1998 {published data only}

Van Lieshout 1998 {published data only}

Verle 1994 {published data only}

Warren 1979 {published data only}

Wilkins 1979 {published data only}

Zumstein 1983 {published data only}