Improved rapid diagnostic tests to detect syphilis and yaws: a systematic review and meta-analysis

Ying Zhang, Su Mei Goh, Maeve B Mello, Rachel C Baggaley, Teodora Wi, Cheryl C Johnson, Kingsley B Asiedu, Michael Marks, Minh D Pham, Christopher K Fairley, Eric P F Chow, Oriol Mitjà, Igor Toskin, Ronald C Ballard, Jason J Ong

ABSTRACT
Background Current rapid tests for syphilis and yaws can detect treponemal and non-treponemal antibodies. We aimed to critically appraise the literature for rapid diagnostic tests (RDTs) which can better distinguish an active infection of syphilis or yaws.

Methods We conducted a systematic review and meta-analysis, searching five databases between January 2010 and October 2021 (with an update in July 2022). A generalised linear mixed model was used to conduct a bivariate meta-analysis for the pooled sensitivity and specificity. Heterogeneity was assessed using the I² statistic. We used the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) to assess the risk of bias and Grading of Recommendations, Assessment, Development and Evaluations (GRADE) to evaluate the certainty of evidence.

Results We included 17 studies for meta-analyses. For syphilis, the pooled sensitivity and specificity of the treponemal component were 0.93 (95% CI: 0.86 to 0.97) and 0.98 (95% CI: 0.96 to 0.99), respectively. For the non-treponemal component, the pooled sensitivity and specificity were 0.90 (95% CI: 0.82 to 0.95) and 0.97 (95% CI: 0.92 to 0.99), respectively. For yaws, the pooled sensitivity and specificity of the treponemal component were 0.86 (95% CI: 0.66 to 0.95) and 0.97 (95% CI: 0.94 to 0.99), respectively. For the non-treponemal component, the pooled sensitivity and specificity were 0.80 (95% CI: 0.55 to 0.93) and 0.96 (95% CI: 0.92 to 0.98), respectively.

Conclusions RDTs that can differentiate between active and previously treated infections could optimise management by providing same-day treatment and reducing unnecessary treatment.

KEY MESSAGE
We systematically reviewed the performance characteristics and clinical utility of rapid diagnostic tests (RDTs) for syphilis and yaws. We report a slightly lower sensitivity, but very high specificity compared with laboratory reference tests. RDTs could reduce time-to-treatment, over-treatment and lost-to-follow-up.

BACKGROUND
Syphilis and yaws are human treponematoses that remain significant causes of morbidity and mortality globally. Syphilis is caused by Treponema pallidum subspecies pallidum and is primarily transmitted through sex by skin-to-skin contact or through mother-to-child during pregnancy, causing congenital syphilis. Yaws is an endemic and neglected tropical disease caused by Treponema pallidum subspecies pertenue and is characterised by soft tissue and bone lesions. Both infections are curable and preventable.

Globally, there are an estimated 6 million new cases of syphilis each year. The burden of congenital syphilis is high, with an estimated 661,000 cases. Further, syphilis disproportionately affects key populations such as sex workers, transgender women (TGW) and men who have sex with men (MSM). Recently, a 2021 study estimated a pooled prevalence of 7.5% among MSM worldwide. Social and structural challenges often make it difficult for these populations to access healthcare services, resulting in delayed detection and lost to follow-up (LTFU) (from diagnosis to getting results or treatment).

For yaws, a systematic review in 2015 estimated the prevalence of active disease ranged from 0.3% to 14.5% in endemic areas, and of latent yaws from 2.5% to 31.1%. Considering its severe morbidity, the WHO launched a strategy to eradicate yaws by 2020, later revised to 2030. The revised strategy included using rapid diagnostic tests (RDTs) for T. pallidum as a priority for yaws eradication.

Diagnostic methods for active syphilis and yaws include direct detection of treponemes or treponemal DNA sequences (ie, darkfield microscopy, direct immunofluorescence test or nucleic acid amplification tests performed on material obtained from primary or secondary lesions). In the absence of primary or secondary lesions, such as in latent syphilis or tertiary syphilis, serological tests for treponemal and non-treponemal antibodies using whole blood, serum/plasma or cerebrospinal fluid are required. Over the past decade, several treponemal rapid screening tests have been developed with pooled sensitivity ranging from 85% to 98%, and specificity from 93% to 98%. In 2015, syphilis RDTs were adopted into the WHO prequalification system. However, these single-treponemal RDTs cannot differentiate between active and previously treated infections.

More recently, some novel RDTs have included both treponemal and non-treponemal test
components in the same device, such as the Dual Path Platform (DPP) Syphilis Screen and Confirm Assay (Chembio Diagnostic Systems, New York, USA), which will be referred to as the DPP-RDT. The Burnet Institute (Melbourne, Australia) also developed an RDT for syphilis using a treponemal IgA-specific assay. Furthermore, a new smartphone dongle triplex test targeting HIV, treponemal antibodies and anti-cardiolipin antibodies as the non-treponemal marker has been developed. Of these novel RDTs, the only commercially available test currently is the DPP Screen and Confirm Assay which is accessible in Europe and the USA. The smartphone dongle and the Burnet tests are prototypes only at this stage and not yet commercially manufactured.

In 2016, a meta-analysis on DPP-RDT to detect syphilis and yaws found an 85.2% concordance when comparing the DPP-RDT with reference serology. Since that publication, there have been further studies evaluating DPP-RDT in various settings, including the use of digital readers as well as newer RDTs. Thus, we conducted a systematic review on the performance characteristics and clinical utility of RDTs for syphilis and yaws to inform forthcoming WHO guidance on testing for these diseases.

METHODS
This review follows the recommendations in the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) extension for Diagnostic Test Accuracy guidelines.

Search strategy and selection criteria
Five databases (Medline, Embase, Global Health, CINAHL and Web of Science) were searched on 11 October 2021. The search strategy was adapted from a previous meta-analysis paper on DPP-RDT, built around overarching terms, including ‘syphilis’, ‘yaws’, ‘rapid diagnostic test’, ‘treponemal’, ‘nontreponemal’ and their Medical Subject Headings (MeSH) terms (eg, syphilis congenital, syphilis latent, neurosyphilis), and was modified for each database (see online supplemental appendix 1). The search was limited from 2010 to October 2021, the period since the DPP-RDT assay became available. No language restrictions were set. Reference lists were checked to locate any other relevant papers.

Studies were included for meta-analysis if they contained primary quantitative data on the clinical performance of an RDT that detects treponemal and non-treponemal antibodies with no restrictions on populations, countries or study designs. Studies that evaluated secondary outcomes such as feasibility, usability and acceptability of the RDTs, testing uptake and cost-effectiveness were included for narrative synthesis.

Search results from each database were downloaded into the Covidence systematic review tool. After removing duplicates, two reviewers (YZ, SMG) independently screened the titles and abstracts of all articles potentially eligible for full-text retrieval, with a third reviewer (JJO) resolving any discrepancies. Non-English language articles were excluded at this stage. An updated literature search was undertaken in July 2022, and after screening, none of the articles met the inclusion criteria.

Data extraction
Two independent reviewers (YZ, SMG) extracted data from full-text articles that satisfied the inclusion criteria using a data extraction spreadsheet and checked by a third reviewer (JJO). We extracted data on the specimen type, disease (syphilis or yaws), RDT reading method (visual vs digital reader), type of laboratory-based reference test, sensitivity and specificity, country (classified by country income level as per the World Bank Group), study design, study setting and secondary outcomes (if available).

Data analysis
The Quality Assessment of Diagnostic Accuracy Studies (QUADAS)-2 tool was chosen to assess the risk of bias in diagnostic test accuracy. Two reviewers (YZ, SMG) examined the risk of bias, and a third reviewer (JJO) resolved any discrepancies. Risk of bias was assessed using QUADAS and the certainty of the evidence was evaluated using Grading of Recommendations, Assessment, Development and Evaluations (GRADE). Statistical analyses were conducted in Stata V.17 (StataCorp, College Station, Texas, USA) using the midas and metandi Stata modules for meta-analysis of diagnostic test accuracy studies. We used a generalised linear mixed-model approach to conduct a bivariate meta-analysis of the sensitivity and specificity of the dual treponemal–non-treponemal RDTs. As the bivariate model assumes independent binomial distributions for the true positives and true negatives conditional on the sensitivity and specificity in each study, we calculated their associated 95% CI. Forest plots and hierarchical summary receiver operating characteristic (HSROC) plots were created separately for syphilis and yaws. Statistical heterogeneity was assessed using the I² statistic, and Deeks’ funnel asymmetry test was used to evaluate publication bias. We calculated the positive and negative likelihood ratios. We used random-effects meta-regression to determine if any study-level covariates could explain the between-study heterogeneity. Meta-regression was not performed for yaws due to insufficient observations. We used narrative synthesis to describe the data for the secondary outcomes. This study is registered with PROSPERO (CRD42021279587).

RESULTS
The quality assessment results are summarised in online supplemental table 1 and online supplemental figure 1. There was potential for bias, particularly for client selection (33.3%, n=5). Most studies adequately described the index test (73.3%, n=11) and reference tests (73.3%, n=11). Results around the certainty of the evidence are shown in online supplemental table 2. The PRISMA flow chart is presented in figure 1. In total, 750 records were retrieved and screened. We included 25 studies for data synthesis, 2 of which were unpublished data extracted from a previous publication following consultation with one of the coauthors (MM). Characteristics of the 25 studies are outlined in table 1. The majority of the studies were cross-sectional studies from high-income countries and conducted within clinical settings. In total, 13 articles on syphilis and 4 articles on yaws included 2 unpublished studies) were included in the meta-analysis (see table 2). Fifteen articles contained enough information for the narrative synthesis of secondary outcomes.

Syphilis
For syphilis, the pooled sensitivity and specificity of the treponemal component were 0.93 (95% CI: 0.86 to 0.97) and 0.98 (95% CI: 0.96 to 0.99), respectively (figure 2). For the non-treponemal component, the pooled sensitivity and specificity for syphilis were 0.90 (95% CI: 0.82 to 0.95) and 0.97 (95% CI: 0.92 to 0.99), respectively (figure 3). High heterogeneity was observed for both the treponemal (sensitivity: I²=96.9%; specificity: I²=95.0%) and non-treponemal (sensitivity: I²=98.6%; specificity: I²=96.5%) components.
specification: $I^2=94.7\%$) and non-treponemal (sensitivity: $I^2=98.3\%$; specificity: $I^2=99.3\%$) components. From the bivariate analysis, the positive and negative likelihood ratios were 55.1 (95% CI: 26.6 to 113.9) and 0.07 (95% CI: 0.04 to 0.14), respectively, for the treponemal component and 34.7 (95% CI: 11.4 to 106.1) and 0.10 (95% CI: 0.06 to 0.18) for the non-treponemal component. The diagnostic ORs were 777 (95% CI: 340 to 1776) and 339 (95% CI: 131 to 880), respectively.

Meta-regression was conducted using the study setting, sample type and RDT reading method (see online supplemental table 3). Serum samples performed better than whole blood samples in both treponemal (0.96 (95% CI: 0.93 to 1.00) vs 0.88 (95% CI: 0.79 to 0.97)) and non-treponemal sensitivity (0.95 (95% CI: 0.92 to 0.99) vs 0.83 (95% CI: 0.70 to 0.91)), but not for specificity. Studies conducted in laboratories had better sensitivity for both treponemal (0.95 (95% CI: 0.83 to 1.00)) and non-treponemal (0.93 (95% CI: 0.86 to 0.99)) test components compared with studies from clinical facilities (0.91 (95% CI: 0.82 to 1.00); 0.85 (95% CI: 0.72 to 0.98)). Although the use of digital readers to analyse RDT results resulted in greater specificity than the human eye (treponemal: 0.99 (95% CI: 0.99 to 1.00) vs 0.98 (95% CI: 0.96 to 0.99); non-treponemal: 0.99 (95% CI: 0.92 to 1.00) vs 0.97 (95% CI: 0.93 to 1.00), respectively), it only had slightly better sensitivity for the treponemal component (0.95 (95% CI: 0.86 to 1.00) vs 0.92 (95% CI: 0.87 to 0.98)) and added to the cost of the test.

Among all the studies, there were two outlier studies that were performed in clinical settings. A study in the USA reported logistical problems with expired test facilities (0.91 (95% CI: 0.82 to 1.00)) and added to the cost of the test. In another outlier study exploring point-of-care tests for syphilis among MSM in Italy, Zorzi et al reported logistical problems with expired test assays that resulted in a subsample of recruited MSM’s results being unavailable. These might have contributed to the heterogeneity in the pooled studies. Further sensitivity analysis by removing these two studies increased the pooled sensitivity from 0.83 (95% CI: 0.82 to 0.95) to 0.98 (95% CI: 0.96 to 0.99) for the non-treponemal part. The $I^2$ statistic was reduced from 96.9% to 87.5%, and 98.3% to 98.2% for the sensitivity of the treponemal and non-treponemal components, respectively. The exclusion of these two studies increased the test performance of the non-treponemal component in clinical settings to 0.92 (95% CI: 0.85 to 1.00).

**Yaws**
For yaws, we found that for the treponemal component, the pooled sensitivity and specificity were 0.86 (95% CI: 0.66 to 0.95) and 0.97 (95% CI: 0.94 to 0.99), respectively, and for the non-treponemal component, 0.80 (95% CI: 0.55 to 0.93) and 0.96 (95% CI: 0.92 to 0.98), respectively (figures 4 and 5). The $I^2$ for sensitivity was 96.4% and 97.8%, and that for specificity was 84.2% and 88.5% for treponemal and non-treponemal components, respectively. The HSROCs for syphilis and yaws are depicted in online supplemental figure 2.

The positive and negative likelihood ratios were 27.8 (95% CI: 12.3 to 63.0) and 0.15 (95% CI: 0.06 to 0.39), respectively, for the treponemal component and 21.8 (95% CI: 8.9 to 53.5) and 0.21 (95% CI: 0.08 to 0.54) for the non-treponemal component. The diagnostic ORs were 187 (95% CI: 39 to 901) and 105 (95% CI: 20 to 553), respectively. Using Deeks’ test, we did not detect any publication bias in the studies on syphilis (treponemal component: p=0.74; non-treponemal component: p=0.53) and yaws (treponemal component: p=0.74; non-treponemal component: p=0.70) (see online supplemental figure 3). The positive predictive values and negative predictive values for tests undertaken for syphilis and yaws are presented in online supplemental table 4.

**Secondary outcomes**
The narrative synthesis of the secondary outcomes is provided in online supplemental appendix 2. Briefly, RDTs were considered acceptable and feasible by healthcare workers and clients, and could reduce time to treatment, LTFU, overtreatment and improve cost-effectiveness. The usability of DPP-RDT was variable, with some studies advocating for digital readers to improve test accuracy.
DISCUSSION

This systematic review synthesised current evidence regarding RDTs for detecting both treponemal and non-treponemal antibodies for syphilis and yaws. Since the last review by Marks et al., new studies have evaluated DPP-RDT in various settings, and two new studies have data on the Burnet assay. We consolidated evidence regarding the acceptability, feasibility, usability, cost-effectiveness and uptake of treatment post-diagnosis, providing helpful information for policy and planning (see online supplemental appendix 2).

Syphilis

While we observed high pooled sensitivity and specificity in our results, we acknowledge that it is challenging to define active syphilis using diagnostics without further medical history (including past syphilis results) and clinical examination (for signs of syphilis). In addition, no test will be 100% accurate and have limitations. According to Shields’s study, routine PCR has a sensitivity of 84–89% and a specificity of 93–100% for primary syphilis, but sensitivity dropped to 50% for secondary syphilis, rendering it unsuitable as a screening tool for secondary syphilis. Other studies report that although venereal disease research laboratory (VDRL) is specific for syphilis, it is more prone to human error and lacks the sensitivity to be used as a first-line screening test for primary syphilis. Serum RPR and VDRL have 62–100% sensitivity, depending on the disease stage. Although we could not stratify our results by different syphilis stages, our results demonstrated strong test performance even with a mix of disease stages.

Notably, we found that serum samples performed better than whole blood samples in test sensitivity but not for specificity. This finding is concordant with Jafari et al, where diagnostic performance for serum samples was higher than whole blood due to higher concentration of biomarkers and absence of interfering substances in whole blood. In addition, we found higher test sensitivity in studies performed in laboratory settings than in clinic settings. This opens the possibility of using highly sensitive RDT for serum samples in laboratory settings, especially in antenatal syphilis screening, where no cases should be missed for treatment. On the other hand, the lower sensitivity of RDTs in the field may be an acceptable trade-off if RDTs can improve detection and reduce LTFU.

Early testing and treatment for syphilis are critical for pregnant women to prevent congenital disease and other negative pregnancy outcomes. Scaling up the use of these newer dual treponemal–non-treponemal RDTs for syphilis could potentially benefit pregnant women and their babies. A modelling study comparing dual RDT with laboratory RPR+T. pallidum haemagglutination (TPHA) estimated that with every 1000 pregnancies, 34 and 26 adverse pregnancy outcomes would be averted, respectively with dual RDT versus RPR+TPHA. Additionally, when RPR+TPHA was used to diagnose maternal syphilis, treatment rates declined from 100% to 67%, indicating that a significant number of clients were LTFU. Hence, the WHO recommends immediate treatment initiation following any reactive syphilis test for pregnant women and their partner(s). This strategy may result in overtreatment due to false positives for previous syphilis infections, it is preferred to avoid missing syphilis treatment during pregnancy. The ability of the RDT to obtain results and initiate treatment at the same antenatal visit can reduce LTFU, prevent more cases of adverse birth outcomes and interrupt the chain of transmission, thus saving valuable client and provider time and resources.

Priority populations such as MSM and TGW are disproportionately affected by syphilis, and the presence of sociocultural stigma, violence, negative experiences with healthcare systems, prioritisation of hormone therapy by transgender people and frequent life instability place them at a higher risk of LTFU. While we observed high pooled sensitivity and specificity in our results, we acknowledge that it is challenging to define active syphilis using diagnostics without further medical history (including past syphilis results) and clinical examination (for signs of syphilis). 

Table 1 Characteristics of included studies

<table>
<thead>
<tr>
<th>POCTs</th>
<th>Syphilis* (n=19) (%)</th>
<th>Yaws* (n=7) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chembio DPP-RDT</td>
<td>15 (78.9)</td>
<td>7 (100)</td>
</tr>
<tr>
<td>Smartphone dongle triplex test†</td>
<td>1 (5.3)</td>
<td>0</td>
</tr>
<tr>
<td>SpanDiagnostics‡</td>
<td>1 (5.3)</td>
<td>0</td>
</tr>
<tr>
<td>Burnet TP-IgA</td>
<td>2 (10.5)</td>
<td>0</td>
</tr>
<tr>
<td>Country income level§</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>9 (47.4)</td>
<td>0</td>
</tr>
<tr>
<td>Middle</td>
<td>6 (31.6)</td>
<td>6 (85.7)</td>
</tr>
<tr>
<td>Low</td>
<td>2 (10.5)</td>
<td>1 (14.3)</td>
</tr>
<tr>
<td>Mixed</td>
<td>2 (10.5)</td>
<td>0</td>
</tr>
<tr>
<td>Study setting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General practice/clinic</td>
<td>11 (57.9)</td>
<td>1 (14.3)</td>
</tr>
<tr>
<td>Laboratory</td>
<td>8 (42.1)</td>
<td>0</td>
</tr>
<tr>
<td>Field/non-clinical facility</td>
<td>0</td>
<td>3 (42.9)</td>
</tr>
<tr>
<td>Unclear (includes unpublished data)</td>
<td>0</td>
<td>3 (42.9)</td>
</tr>
<tr>
<td>Population</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General population</td>
<td>11 (57.9)</td>
<td>3 (42.9)</td>
</tr>
<tr>
<td>Pregnant women</td>
<td>4 (21.1)</td>
<td>0</td>
</tr>
<tr>
<td>MSM</td>
<td>2 (10.5)</td>
<td>0</td>
</tr>
<tr>
<td>People living with HIV</td>
<td>1 (5.3)</td>
<td>0</td>
</tr>
<tr>
<td>Children</td>
<td>0</td>
<td>2 (28.6)</td>
</tr>
<tr>
<td>Other¶</td>
<td>1 (5.3)</td>
<td>0</td>
</tr>
<tr>
<td>Unclear (from unpublished data)</td>
<td>0</td>
<td>2 (28.6)</td>
</tr>
<tr>
<td>Study design</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental/randomised controlled trial</td>
<td>0</td>
<td>1 (14.3)</td>
</tr>
<tr>
<td>Observational/cross-sectional</td>
<td>16 (84.2)</td>
<td>2 (28.6)</td>
</tr>
<tr>
<td>Modelling</td>
<td>2 (10.5)</td>
<td>1 (14.3)</td>
</tr>
<tr>
<td>Qualitative</td>
<td>1 (5.3)</td>
<td>1 (14.3)</td>
</tr>
<tr>
<td>Unclear (from unpublished data)</td>
<td>0</td>
<td>2 (28.9)</td>
</tr>
<tr>
<td>RDT reading method</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual</td>
<td>16 (84.2)</td>
<td>6 (85.7)</td>
</tr>
<tr>
<td>Digital reader</td>
<td>3 (15.8)</td>
<td>1 (14.3)</td>
</tr>
<tr>
<td>Secondary outcomes</td>
<td>Total (n=15)</td>
<td></td>
</tr>
<tr>
<td>Acceptability</td>
<td>2 (13)</td>
<td></td>
</tr>
<tr>
<td>Feasibility</td>
<td>2 (13)</td>
<td></td>
</tr>
<tr>
<td>Usability</td>
<td>5 (33)</td>
<td></td>
</tr>
<tr>
<td>Appropriate treatment following testing</td>
<td>4 (27)</td>
<td></td>
</tr>
<tr>
<td>Cost/resources</td>
<td>2 (13)</td>
<td></td>
</tr>
</tbody>
</table>

*The total of studies for each category does not add up to 25 as one paper contained data for both syphilis and yaws.
†HIV, treponemal and non-treponemal RDT, not commercially available.
‡Similar to DPP-RDT, manufactured by Span Diagnostics.
§Country income level is classified as per the World Bank Group.
¶(>15 years old)+behavioural risk group: (1) injection drug users (IDUs) with verified track marks (eg, visible signs of injection); (2) women who reported at least two male partners in the last 2 years or engaging in anal intercourse, sex trading, or sex with an MSM, an IDU, or an HIV-positive man; (3) MSM and men who have sex with men and women; and (4) transgender individuals.
DPP, Dual Path Platform; MSM, men who have sex with men; POCTs, point-of-care tests; RDT, rapid diagnostic test; TP-IgA, treponemal IgA-specific assay.
### Table 2  Summary of studies included in the meta-analysis (n=17)

<table>
<thead>
<tr>
<th>Author</th>
<th>Study site</th>
<th>Population</th>
<th>T1 reference test</th>
<th>T1 reference prevalence (%)</th>
<th>T2 reference test</th>
<th>T2 reference prevalence (%)</th>
<th>Sample type</th>
<th>Sample size</th>
<th>T1 sensitivity (95% CI)</th>
<th>T1 specificity (95% CI)</th>
<th>T2 sensitivity (95% CI)</th>
<th>T2 specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Syphilis (DPP-RDT)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Castro‡</td>
<td>USA General</td>
<td>TPHA</td>
<td>40.2</td>
<td></td>
<td>RPR</td>
<td>30.6</td>
<td>Serum</td>
<td>316</td>
<td>0.97 (0.93 to 0.99)*</td>
<td>0.99 (0.97 to 1.00)*</td>
<td>0.97 (0.91 to 0.99)*</td>
<td>0.98 (0.95 to 0.99)*</td>
</tr>
<tr>
<td>Castro‡</td>
<td>USA General</td>
<td>TPHA</td>
<td>62.9</td>
<td></td>
<td>RPR</td>
<td>52.1</td>
<td>Serum</td>
<td>1601</td>
<td>0.97 (0.95 to 0.98)*</td>
<td>0.95 (0.93 to 0.97)*</td>
<td>0.89 (0.86 to 0.91)*</td>
<td>0.99 (0.97 to 0.99)*</td>
</tr>
<tr>
<td>Castro‡</td>
<td>Portugal General</td>
<td>TPHA</td>
<td>74.6</td>
<td></td>
<td>RPR</td>
<td>69.8</td>
<td>Serum</td>
<td>248</td>
<td>0.99 (0.97 to 1.00)</td>
<td>0.89 (0.78 to 0.95)</td>
<td>0.99 (0.95 to 1.00)</td>
<td>0.95 (0.87 to 0.98)</td>
</tr>
<tr>
<td>Coarse‡</td>
<td>Australia MSM</td>
<td>TPHA</td>
<td>73.2</td>
<td></td>
<td>RPR</td>
<td>55</td>
<td>Serum</td>
<td>1005</td>
<td>0.9 (0.87 to 0.92)</td>
<td>0.99 (0.97 to 1.00)</td>
<td>0.94 (0.92 to 0.96)</td>
<td>0.62 (0.58 to 0.67)</td>
</tr>
<tr>
<td>Constantine‡</td>
<td>USA General</td>
<td>TPHA</td>
<td>31.5</td>
<td></td>
<td>RPR</td>
<td>31.2</td>
<td>Whole blood</td>
<td>1266</td>
<td>0.93 (0.90 to 0.95)</td>
<td>0.99 (0.99 to 1.00)</td>
<td>0.65 (0.60 to 0.70)</td>
<td>1.00 (1.00 to 1.00)</td>
</tr>
<tr>
<td>Guinard‡</td>
<td>France General</td>
<td>BA</td>
<td>57.6</td>
<td></td>
<td>RPR</td>
<td>39.8</td>
<td>Serum whole blood</td>
<td>T1=144</td>
<td>0.9 (0.82 to 0.95)</td>
<td>0.98 (0.91 to 1.00)</td>
<td>0.95 (0.84 to 0.98)</td>
<td>0.92 (0.83 to 0.97)</td>
</tr>
<tr>
<td>Hess‡</td>
<td>USA Others</td>
<td>TPHA</td>
<td>12.2</td>
<td></td>
<td>RPR</td>
<td>3</td>
<td>Whole blood</td>
<td>T1=765</td>
<td>0.53 (0.43 to 0.63)</td>
<td>0.99 (0.97 to 0.99)</td>
<td>0.48 (0.27 to 0.69)</td>
<td>0.99 (0.98 to 1.00)</td>
</tr>
<tr>
<td>Langendorf‡</td>
<td>Burkina Faso Pregnant women</td>
<td>TPHA</td>
<td>41.7</td>
<td></td>
<td>RPR</td>
<td>25.5</td>
<td>Finger prick whole blood</td>
<td>T1=242</td>
<td>0.95 (0.89 to 0.98)*</td>
<td>0.98 (0.94 to 1.00)*</td>
<td>0.85 (0.72 to 0.92)*</td>
<td>1.00 (1.00 to 1.00)*</td>
</tr>
<tr>
<td>Skinner‡</td>
<td>Australia Children</td>
<td>TPHA</td>
<td>75.9</td>
<td></td>
<td>RPR</td>
<td>50.6</td>
<td>Serum</td>
<td>449</td>
<td>0.94 (0.90 to 0.96)*</td>
<td>0.87 (0.79 to 0.93)*</td>
<td>0.96 (0.92 to 0.98)*</td>
<td>0.66 (0.60 to 0.72)*</td>
</tr>
<tr>
<td>Yin‡</td>
<td>China General</td>
<td>TPHA</td>
<td>49.9</td>
<td></td>
<td>TRUST</td>
<td>35.6</td>
<td>Finger prick whole blood</td>
<td>3135</td>
<td>0.96 (0.95 to 0.97)</td>
<td>0.99 (0.99 to 1.00)</td>
<td>0.89 (0.87 to 0.91)</td>
<td>0.91 (0.90 to 0.92)</td>
</tr>
<tr>
<td>Zorzi‡</td>
<td>Italy MSM</td>
<td>TPHA, CLIA</td>
<td>15.4</td>
<td></td>
<td>RPR</td>
<td>7.1</td>
<td>Finger prick whole blood</td>
<td>227</td>
<td>0.69 (0.51 to 0.83)</td>
<td>0.99 (0.96 to 1.00)</td>
<td>0.62 (0.35 to 0.85)</td>
<td>1.00 (1.00 to 1.00)</td>
</tr>
<tr>
<td><strong>Yaws (DPP-RDT)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ayove‡</td>
<td>Papua New Guinea Children</td>
<td>TPHA</td>
<td>55.3</td>
<td></td>
<td>RPR</td>
<td>43.4</td>
<td>Plasma whole blood</td>
<td>704900</td>
<td>0.88 (0.85 to 0.91)</td>
<td>0.95 (0.92 to 0.97)</td>
<td>0.88 (0.84 to 0.91)</td>
<td>0.92 (0.89 to 0.95)</td>
</tr>
<tr>
<td>Aziz‡</td>
<td>Ghana Unknown</td>
<td>TPHA</td>
<td>39.2</td>
<td></td>
<td>RPR</td>
<td>34.9</td>
<td>Finger prick</td>
<td>215</td>
<td>0.97 (0.91 to 0.99)</td>
<td>0.99 (0.96 to 1.00)</td>
<td>0.94 (0.87 to 0.98)</td>
<td>0.99 (0.96 to 1.00)</td>
</tr>
<tr>
<td>Marks‡</td>
<td>Solomon Islands Children</td>
<td>TPHA</td>
<td>29.6</td>
<td></td>
<td>RPR</td>
<td>28.9</td>
<td>Serum</td>
<td>415</td>
<td>0.59 (0.50 to 0.67)</td>
<td>0.98 (0.95 to 0.99)</td>
<td>0.42 (0.33 to 0.51)</td>
<td>0.95 (0.92 to 0.97)</td>
</tr>
<tr>
<td>Taleo‡</td>
<td>Vanuatu Unknown</td>
<td>TPHA</td>
<td>43.7</td>
<td></td>
<td>RPR</td>
<td>345</td>
<td>Finger prick</td>
<td>238</td>
<td>0.8 (0.71 to 0.87)</td>
<td>0.93 (0.88 to 0.96)</td>
<td>0.74 (0.64 to 0.83)</td>
<td>0.96 (0.92 to 0.99)</td>
</tr>
<tr>
<td><strong>Syphilis (Burnet TP-IgA)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pham10</td>
<td>China General</td>
<td>TPHA</td>
<td>33.7</td>
<td></td>
<td>RPR</td>
<td>33.7</td>
<td>Plasma</td>
<td>454</td>
<td>0.96 (0.92 to 0.99)</td>
<td>0.85 (0.80 to 0.89)</td>
<td>0.99 (0.98 to 1.00)</td>
<td>0.99 (0.98 to 1.00)</td>
</tr>
<tr>
<td>Pham11</td>
<td>South Africa Pregnant women</td>
<td>TPAAb</td>
<td>0.6</td>
<td></td>
<td>RPR</td>
<td>0.6</td>
<td>Whole blood</td>
<td>499</td>
<td>0.88 (0.29 to 1.00)*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Upper limit of 95% CI above 0.995 is rounded up to 1.00.
*Values are calculated by authors as they were not reported in the original studies.
†Sub-analysis with RPR titre.
‡Unpublished data.
CLIA, chemiluminescence immunoassay; DPP-RDT, Dual Path Platform-rapid diagnostic test; BA, enzyme immunoassay; MSM, men who have sex with men; RPR, rapid plasma reagin; TPAAb, Treponema pallidum antibody; TPHA, Treponema pallidum haemagglutination; TP-IgA, treponemal IgA-specific assay; TTPA, Treponema pallidum passive particle agglutination assay; TRUST, toluidine red unheated serum test.

for a confirmatory test. Although test performance of RDT is slightly lower in clinical settings than in laboratories, given their high prevalence and LTFU, RDTs could be preferred over conventional laboratory testing. The added value of newer syphilis RDTs, compared with single-treponemal rapid screening tests or conventional laboratory-based testing, lies in facilitating therapy on the same day and reducing overtreatment, particularly among users of HIV pre-exposure prophylaxis and in areas with a high background prevalence of syphilis. Given that they are recommended to undergo syphilis tests every 3–6 months,

### Figure 2
Forest plot of treponemal sensitivity and specificity for syphilis.

### Figure 3
Forest plot of non-treponemal sensitivity and specificity for syphilis.
treatment based solely on a positive single-treponemal rapid test will result in significant overtreatment.

**Yaws**
Access to quality diagnostics has been identified as a priority in controlling, eliminating and eradicating neglected tropical diseases, and the expanded use of RDTs for yaws is central to WHO’s eradication effort. Currently, most countries rely solely on clinical diagnosis, which is not sufficiently accurate and leads to unreliable surveillance data. RDTs allow easier identification of cases of latent yaws in the community who potentially represent an important disease reservoir.42 As most yaws-endemic countries lack sufficient laboratory capacity for traditional serological assays, these novel RDTs play a pivotal role in supporting

Figure 4  Forest plot of treponemal sensitivity and specificity for yaws. *Unpublished studies.

Figure 5  Forest plot of non-treponemal sensitivity and specificity for yaws. *Unpublished studies.
yaws eradication efforts. The use of additional automatic readers can potentially monitor changes in the quantity of the non-treponemal antibodies, thereby assisting in the diagnosis of new infections or monitoring treatment response. In Papua New Guinea, children with yaws were followed up using a DPP-RDT automatic reader to measure optical density after treatment.34 At 6 months, 95% had attained a fourfold reduction in optical density (serological cure) or seroconversion.34 This demonstrates that post-treatment serological follow-up might be done in the same way that reference RPR testing is used without relying on laboratory facilities. In a community surveillance study, Marks et al reported the sensitivity of the DPP-RDT against T. pallidum passive particle agglutination assay and RPR was 47.1%, with the sensitivity of the DPP-RDT being strongly related to the RPR titre. This reduced sensitivity compared with other studies reflects a greater population of asymptomatic latent yaws cases where lower antibody titres contribute to lower sensitivity compared with those with active clinical disease and higher titres.33 This is important, particularly in antenatal settings, as pregnant women with yaws and lower RPR titres may be less likely to transmit the infection to their infants.

Our review has several limitations. First, many studies were performed in a laboratory setting and included samples with different patterns of serological reactivity but unknown clinical stages of infection. Further comparative studies are needed in syphilis and yaws, where the clinical stages of infection are documented together with direct detection of treponemes (in primary and secondary disease), clinical and treatment histories (including information about serofast status) so that active disease can be ascertained with greater certainty. Second, we did not have information on coinfection status, re-infection status or other diseases in subjects providing samples that might have affected the results. Third, we did not search grey literature, so we may have missed other relevant data. Lastly, we tried to use meta-regression to explain the heterogeneity in our results but was limited by the small number of studies and not enough information to account for other important factors such as the clinical stages of syphilis and yaws, and treatment histories of patients.

CONCLUSIONS

RDTs that can differentiate between active and previously treated infections could optimise management by providing same-day treatment and reducing unnecessary treatment. This systematic review and meta-analysis found that current RDTs for syphilis and yaws had slightly lower sensitivity but a very high specificity than laboratory-based testing. If distributed widely with appropriate training, these tests can potentially decrease the incidence of both adult and congenital syphilis and contribute to the global eradication of yaws.

Author affiliations
1School of Public Health, The University of Sydney, Camperdown, New South Wales, Australia
2Melbourne Sexual Health Centre, Melbourne, Victoria, Australia
3Global HIV, Hepatitis and STI Programmes, WHO, Geneva, Switzerland
4Clinical Research Department, London School of Hygiene and Tropical Medicine, London, UK
5Hospital for Tropical Diseases, University College London Hospital, London, UK
6Division of Infection and Immunity, University College London, London, UK
7Burnet Institute, Melbourne, Victoria, Australia
8School of Public Health and Preventive Medicine, Monash University Faculty of Medicine, Nursing and Health Sciences, Melbourne, Victoria, Australia
9Central Clinical School, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Victoria, Australia
10Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia
11Fight AIDS and Infectious Diseases Foundation, Catalonia, Spain
12Department of Sexual and Reproductive Health and Research, WHO, Geneva, Switzerland

Correction notice This article has been corrected since it was first published online. The disclaimer section has been updated.

Handling editor Laih J Abu-Raddad
Twitter Ying Zhang @love_cally, Eric P F Chow @EricPChow and Jason J Ong @DrlassonOng

Contributors MM, TW, RB, CI and JIO conceived the idea. YZ and SMG did the screening and data extraction. YZ and JIO conducted the statistical analysis. All authors contributed to the interpretation of the results and subsequent edits of the manuscript and had final responsibility for the decision to submit for publication.

Funding The WHO supported this work from a grant from the Ministry for Development Cooperation and Humanitarian Affairs, Luxembourg, JIO and EPFC are each supported by the Australian National Health and Medical Research Council (NHMRC) Emerging Leadership Investigator Grant (grant number GNT1139355 for JIO; GNT1172873 for EPFC). CKF is supported by an Australian NHMRC Leadership Investigator Grant (grant number GNT1172900).

Disclaimer Some of the authors are present or former staff members of the World Health Organization. The authors alone are responsible for the views expressed in this publication and they do not necessarily represent the views, decisions, or policies of the institutions with which they are affiliated.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval No ethical clearance was required.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Ying Zhang http://orcid.org/0000-0001-7717-5691
Michael Marks http://orcid.org/0000-0002-7585-4743
Minh D Pham http://orcid.org/0000-0002-5932-3491
Christopher K Fairley http://orcid.org/0000-0001-9081-1664
Eric P F Chow http://orcid.org/0000-0003-1766-0657
Jason J Ong http://orcid.org/0000-0001-5784-7403

REFERENCES