Supplementary appendix 1. Search strategies.

PUBMED search strategy

Search	Query
#17	Search (((((("Syphilis"[Mesh]) OR syphil*)) AND (((((pregnan*) OR maternal) OR gestation)) OR "Pregnancy"[Mesh]))) AND ((((("Diagnosis"[Mesh]) OR ((diagnos*) OR screening))) OR ((((("Serology"[Mesh]) OR "Serologic Tests"[Mesh])) OR serologic*)))) AND ((((("Sensitivity and Specificity"))) OR ((((((sensitivity) OR specificity) OR post-test probability) OR predictive value$) OR likelihood ratio$) OR ((pre-test or pretest) adj probability)))
#16	Search (((((sensitivity) OR specificity) OR post-test probability) OR predictive value$) OR likelihood ratio$) OR ((pre-test or pretest) adj probability)
#15	Search ("Sensitivity and Specificity")
#14	Search (((((("Syphilis"[Mesh]) OR syphil*)) AND (((((pregnan*) OR maternal) OR gestation)) OR "Pregnancy"[Mesh]))) AND ((((("Diagnosis"[Mesh]) OR ((diagnos*) OR screening))) OR ((((("Serology"[Mesh]) OR "Serologic Tests"[Mesh])) OR serologic*))
#13	Search ((("Diagnosis"[Mesh]) OR ((diagnos*) OR screening))) OR ((((("Serology"[Mesh]) OR "Serologic Tests"[Mesh])) OR serologic*)
#12	Search ((("Serology"[Mesh]) OR "Serologic Tests"[Mesh])) OR serologic*
	Search ("Serology"[Mesh]) OR "Serologic Tests"[Mesh]
#10	Search ("Diagnosis"[Mesh]) OR ((diagnos*) OR screening)
#9	Search (diagnos*) OR screening
#8	Search diagnosis[MeSH Terms]
#7	Search (((syphilis[MeSH Terms]) OR syphil*)) AND (((((pregnan*) OR maternal) OR gestation*)) AND pregnancy[MeSH Terms])
#6	Search ((((pregnan*) OR maternal) OR gestation*)) AND pregnancy[MeSH Terms]
#5	Search "Pregnancy"[Mesh]
#4	Search ((pregnan*) OR maternal) OR gestation*
#3	Search (syphilis[MeSH Terms]) OR syphil*
#2	Search syphil*
#1	Search syphilis[MeSH Terms]

EMBASE search strategy

1. exp syphilis/
2. syphil*.mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
3. 1 or 2
4. (pregnan* or maternal* or antenatal* or gestation*).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
5. exp pregnancy/
6. 4 or 5
7. 3 and 6
8. exp diagnosis/
9. exp serology/ or exp serodiagnosis/
10. diagnos*.mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
11. screening.mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
12. serologic*.mp.
13. 8 or 9 or 10 or 11 or 12
14. 7 and 13
15. exp "Sensitivity and Specificity"/
16. (sensitivity or specificity).mp.

17. ((pre-test or pretest) adj probability).mp.

18. (post-test probability or predictive value$ or likelihood ratio$).mp.

19. *Diagnostic Accuracy/

20. 15 or 16 or 17 or 18 or 19

21. 14 and 20

Supplementary appendix 2. Further information on evidence synthesis methods and results.

Statistical model for the meta-analysis

Meta-analysis was undertaken to generate pooled estimates of diagnostic parameters. The number of true positives, false negatives, false positives and true negatives from each study was meta-analysed to estimate sensitivity and specificity under the assumption that the standard reference test (nonTP test followed by TP test where the nonTP test is positive) was 100% sensitive and specific (Jafari *et al*, 2013). In brief, a bivariate normal model was used to model the population logit sensitivities and population logit specificities in each study to account for correlation between sensitivity and specificity within studies (Reitsma *et al,* 2005). It is assumed that the observed number of true positives in study i, TP_i, is binomially distributed with parameter, π_{Ai}, representing the study-specific sensitivity given the total number of positives on the reference test such that:

$$TP_i \sim Binomial\left(\pi_{Ai}, (TP_i + FN_i)\right)$$

Similarly, the observed number of true negatives in study i, TN_i, is assumed to be binomially distributed with parameter, π_{Bi}, representing the study-specific specificity given the total number of negatives on the reference test such that:

$$TN_i \sim Binomial\left(\pi_{Bi}, (FP_i + TN_i)\right)$$

The parameters are transformed to the real line using the logit transformation such that:

$$\mu_{Ai} = logit(\pi_{Ai})$$
$$\mu_{Bi} = logit(\pi_{Bi})$$

Sensitivity and specificity are correlated within study such that higher values for sensitivity tend to be associated with lower values for specificity, and vice versa. This is modelled by assuming that the study-specific logits for sensitivity and specificity arise from a bivariate normal distribution with population logits for sensitivity and specificity, $(\mu_A, \mu_B)^T$, respectively and variance-covariance matrix, Σ_{AB}, such that:

$$\begin{pmatrix}\mu_{Ai}\\ \mu_{Bi}\end{pmatrix} \sim N\left(\begin{pmatrix}\mu_A\\ \mu_B\end{pmatrix}, \Sigma_{\mathrm{AB}}\right)$$

$$\Sigma_{\mathrm{AB}} = \begin{pmatrix}\sigma_{\mathrm{A}}^2 & \sigma_{\mathrm{AB}}\\ \sigma_{\mathrm{AB}} & \sigma_{\mathrm{B}}^2\end{pmatrix}$$

where σ_{A}^2 represents the variability in the logit sensitivities between studies, σ_{B}^2 represents the variability in the logit specificities between studies and σ_{AB} represents the covariance of the logit sensitivity and logit specificity.

The model was completed by giving the uncertain parameters the following reference prior distributions:

- $\mu_A \sim N(0, 1000)$
- $\mu_B \sim N(0, 1000)$

The number of studies included studies was small and so there was insufficient evidence to estimate the between study heterogeneity using the sample data alone. A weakly informative prior was therefore used for the variance-covariance matrix:

- $\Sigma_{\mathrm{AB}} \sim \mathrm{IW}\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \upsilon = 5\right)$

  IW represents the inverse Wishart distribution on υ degrees of freedom. This prior distribution has a between study standard deviation of 0.5 (95% CrI: 0.3, 1.4).

Analyses were conducted in R (R Core Team, 2014) using the WinBUGS (Lunn *et al*, 2000) and R2WinBUGS software packages (Sturtz *et al*, 2005). Convergence to the target posterior distributions was assessed using the Gelman-Rubin convergence statistic (Brooks *et al*, 1998). A burn-in of 10,000 iterations of the Markov chain was used with a further 10,000 iterations retained to estimate parameters. There was no evidence of high autocorrelation between successive samples of the Markov chains.

Results were displayed as forest plots and scatter plots with 95% credible intervals (CrI) and 95% prediction intervals (PrI) for sensitivity and specificity.

Results of the meta-analysis

The between-study standard deviations for sensitivity and specificity were estimated to be 0.58 (95% CrI: 0.33, 1.15) and 0.71 (95% CrI: 0.42, 1.47), respectively. The correlation between logit sensitivity and specificity was 0.42 (95% CrI: -0.46, 0.88).

Supplementary appendix 3. Studies excluded at full-text with reason for exclusion.

AUTHOR, YEAR	Reason for Exclusion
Angue Y *et al*, 2005	Incomplete ST: Only VDRL
Arana FE *et al*, 2015	Incomplete ST: Only TPHA
Bocoum FY *et al*, 2015	Incomplete ST: Only TPHA
Bristow CC *et al*, 2015	Other target condition: HIV/Syphilis Duo
Bristow CC *et al*, 2014	Incomplete ST: Only TPHA
Chiappe MA *et al*, 2013	Other target condition: HIV/Syphilis Duo
Dzokoto AW *et al*, 2013	Insufficient data 2x2 table
Fakile Y *et al*, 2015	Other target condition: HIV/Syphilis Duo
Hernandez-Trejo M *et al*, 2006	Not correct population: puerperal women
Jafari Y *et al*, 2013	Not accuracy study (Systematic Review)
Juárez-Figueroa L *et al*, 2007	Not correct population: puerperal women
Lee J.-H. *et al*, 2015	Incomplete ST: Only TPHA
Loeffelholz MJ *et al*, 2011	Set of different cutoffs
Mabey D *et al*, 2006	Incomplete ST: Only TPPA
Mehra B *et al*, 2016	Incomplete ST: Only TPHA
Nyamwamu LB, 2009	Incomplete ST: Only VDRL or TPHA
Omoding D *et al*, 2014	Incomplete ST: Only TPHA
Rogozinska E et al, 2016	Not accuracy study (Systematic Review)
Shimelis T, Tadesse E, 2015	Other target condition: HIV/Syphilis Duo
Smit PW *et al*, 2013	Incomplete ST: Only TPPA
Tucker JD *et al*, 2010	Not accuracy study (Systematic Review)
Villazón-Vargas N *et al*, 2009	Not correct population: puerperal women
Yin Y.-P *et al*, 2015	Other target condition: HIV/Syphilis Duo

AUTHOR, YEAR	Reason for Exclusion
ST, Standard Test; RPR; TPHA, T pallidum haemagglutination test; TPPA; Treponema pallidum particle agglutination; VDRL, venereal disease research laboratory	

Supplementary appendix 4. Characteristics of included studies.

Author, year Country, setting Study design	Inclusion criteria	No. recruited	Stage of pregnancy	Definition of syphillis	ICS test	ICS samplng method	nonTP, TP test	nonTP, TP samplng method
Benzaken *et al*, 2011 Brazil, 12 antenatal clinics, rural sites Prospective single-gate, consecutive recruitment	Pregnant women attending antenatal clinics with no age restrictions	712	Mean gestational age was 22.4 weeks (SD ±9.3). Equivalent to 3th trimester	Women with positive FTA-Abs and VDRL. High (HTS) and low-titer active syphilis with VDRL titers ≥ or < 1:8, respectively	VisiTect Syphilis test	Was collected a 20-µL fingerprick capillary blood sample for on-site testing	FTA-Abs, VDRL	An 8-mL venous blood sample was drawn and serum was stored at -20C until shipment to the FUAM reference laboratory.
Bronzan *et al*, 2007 South Africa, 8 antenatal clinics, rural sites Prospective single-gate, consecutive recruitment	Pregnant women ≥18 years without a prior syphilis test during current pregnancy seeking antenatal services at clinics offering onsite syphilis screening.	695	Mean 26.6 weeks (median, 28), w), n=79 from intervention clinics with active syphilis	Women with positive TPHA and RPR. High (HTS) and low-titer active syphilis with RPR titers ≥ or < 1:8, respectively	Determine Syphilis TP	Fingerstick blood was obtained for on-site testing	TPHA, RPR	All clients had venous blood drawn for testing in reference laboratory. All sera were tested by quantitative RPR, ICS, and TPHA

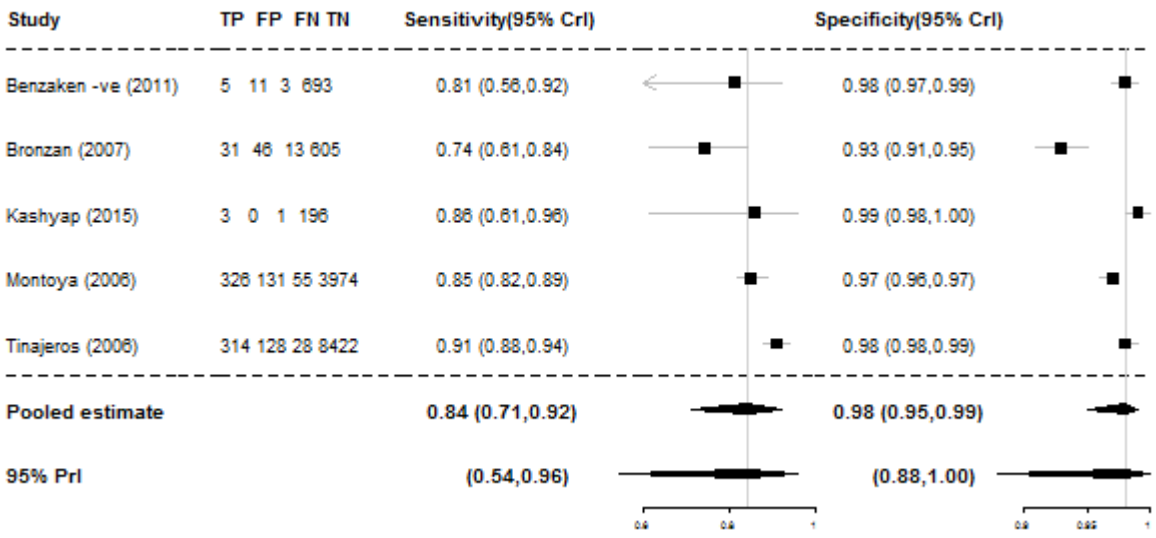
Author, year Country, setting Study design	Inclusion criteria	No. recruited	Stage of pregnancy	Definition of syphillis	ICS test	ICS samplng method	nonTP, TP test	nonTP, TP samplng method
Kashyap *et al*, 2015 India, 1 hospital, urban site Prospective single-gate, recruitment unclear	Pregnant women aged 20-30 years on their first antenatal visit or follow-up visits. Women with previous history of syphilis or any contact history were excluded.	200	1st trimester, 16% 2nd trimester, 32% 3rd trimester, 52%	NR	SD Bioline Syphilis 3.0	Sera separated from the blood samples of all the cases were stored at 4°C till further processing.	VDRL, TPHA	Performance of ICS test was compared with VDRL and TPHA test on each serum sample.
Montoya *et al*, 2006 Mozambique, 6 antenatal clinics, rural sites Prospective single-gate, consecutive recruitment	Pregnant women attending their first antenatal visit	4486	Gestational age at first visit 20.6 (SD ±6.1) weeks	Syphilis status was defined by a gold standard TPHA and RPRRef, and a direct immunofluorescence stain.	SD Bioline Syphilis 3.0	Were collected 20 µL of capillary blood with EDTA or non-EDTA capillary tubes from a finger puncture for ICS testing	RPR, TPHA	Serum from a 5 ml venous blood sample was used for the RPR at the health facilities and the ICS, RPR and TPHA tests at Hospital reference laboratory.

Author, year Country, setting Study design	Inclusion criteria	No. recruited	Stage of pregnancy	Definition of syphillis	ICS test	ICS samplng method	nonTP, TP test	nonTP, TP samplng method
Tinajeros *et al*, 2006 Bolivia, 4 hospitals; large urban sites Prospective single-gate, consecutive recruitment	Pregnant women ≥18 years seeking care at a maternity hospital not tested for syphilis previously during the current pregnancy	8892	NR	Women positive RPR performed by the reference laboratory confirmed by positive TPPA performed by the same laboratory	Determine Syphilis TP	A finger-stick blood sample was obtained using an autolancet and put directly a drop of blood from the finger on ICS test.	RPR, TPPA	Collected a venous puncture blood sample and extracted serum to perform RPR (RPRHosp) on-site, and a separate aliquot was frozen and sent to the reference laboratory to perform another RPR (RPRRef) and confirmed positive RPRRef findings by TPPA

FTA-Abs, fluorescent treponemal antibody absorbed; RPR, rapid plasma reagin; SD, standard deviation; TPHA, T pallidum haemagglutination test; TPPA; Treponema pallidum particle agglutination; VDRL, venereal disease research laboratory

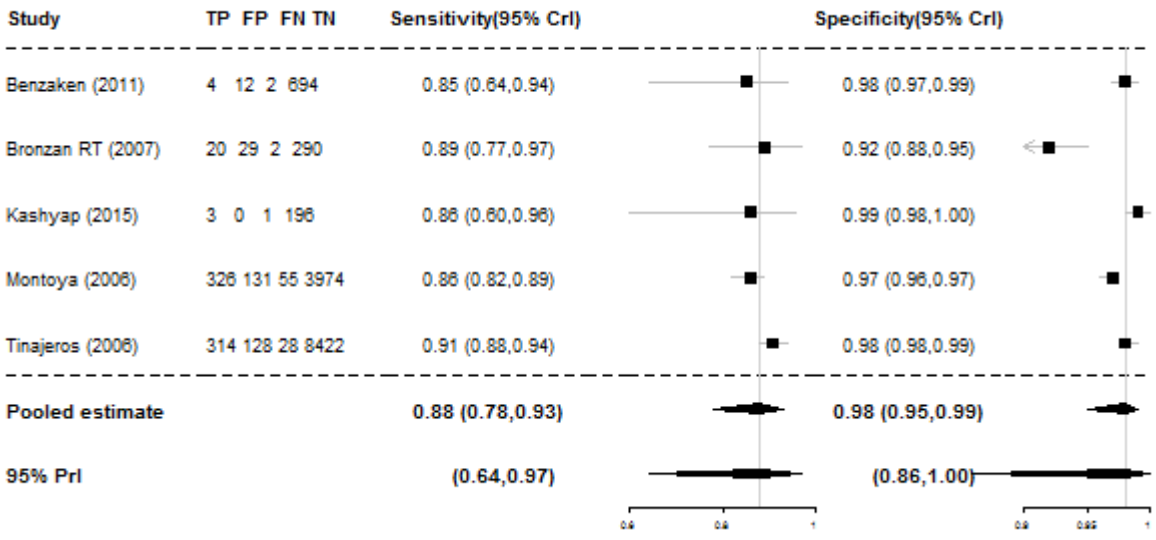
Supplementary appendix 5. Evidence synthesis sensitivity analyses.

Results of the meta-analysis assuming that missing test results in Benzaken et al were negative (rather than positive) are shown in Supplementary Figure 1.



Supplementary Figure 1 Sensitivity and specificity of point-of-care ICS compared with a standard reference test. Sensitivity analysis assuming missing VDRL test negative.

The impact of retraining was assessed by including the data after training only from Bronzan et al. The results are shown in Supplementary Figure 2.



Supplementary Figure 2: Sensitivity and specificity of point-of-care ICS compared with a standard reference test. Sensitivity analysis using results from Bronzan et al after retraining