PUBLISHED VERSION

Verhaegen, J.; Gallos, I.; van Mello, N.; Abdel-Aziz, M.; Takwoingi, Y.; Harb, H.; Deeks, J.; Mol, B.; Coomarasamy, A. Accuracy of single progesterone test to predict early pregnancy outcome in women with pain or bleeding: Meta-analysis of cohort studies BMJ, 2012; 345(Sept 27):e6077

http://hdl.handle.net/2440/84879

RESEARCH

Accuracy of single progesterone test to predict early pregnancy outcome in women with pain or bleeding: meta-analysis of cohort studies

OPEN ACCESS

Jorine Verhaegen *medical student*[1], Ioannis D Gallos *specialist registrar*[2], Norah M van Mello *research fellow*[3], Mohamed Abdel-Aziz *consultant*[2], Yemisi Takwoingi *research fellow*[3], Hoda Harb *research fellow*[2], Jonathan J Deeks *professor*[3], Ben W J Mol *professor*[1], Arri Coomarasamy *professor*[2]

[1]Department of Obstetrics and Gynaecology, Academic Medical Centre University of Amsterdam, Amsterdam, Netherlands; [2]School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK; [3]Public Health, Epidemiology and Biostatistics, School of Health and Population Sciences, College of Medical and Dental Sciences, University of Birmingham

Abstract

Objective To determine the accuracy with which a single progesterone measurement in early pregnancy discriminates between viable and non-viable pregnancy.

Design Systematic review and meta-analysis of diagnostic accuracy studies.

Data sources Medline, Embase, CINAHL, Web of Science, ProQuest, Conference Proceedings Citation Index, and the Cochrane Library from inception until April 2012, plus reference lists of relevant studies.

Study selection Studies were selected on the basis of participants (women with spontaneous pregnancy of less than 14 weeks of gestation); test (single serum progesterone measurement); outcome (viable intrauterine pregnancy, miscarriage, or ectopic pregnancy) diagnosed on the basis of combinations of pregnancy test, ultrasound scan, laparoscopy, and histological examination; design (cohort studies of test accuracy); and sufficient data being reported.

Results 26 cohort studies, including 9436 pregnant women, were included, consisting of 7 studies in women with symptoms and inconclusive ultrasound assessment and 19 studies in women with symptoms alone. Among women with symptoms and inconclusive ultrasound assessments, the progesterone test (5 studies with 1998 participants and cut-off values from 3.2 to 6 ng/mL) predicted a non-viable pregnancy with pooled sensitivity of 74.6% (95% confidence interval 50.6% to 89.4%), specificity of 98.4% (90.9% to 99.7%), positive likelihood ratio of 45 (7.1 to 289), and negative likelihood ratio of 0.26 (0.12 to 0.57). The median prevalence of a non-viable pregnancy was 73.2%, and the probability of a non-viable pregnancy was raised to 99.2% if the progesterone was low. For women with symptoms alone, the progesterone test had a higher specificity when a threshold of 10 ng/mL was used (9 studies with 4689 participants) and predicted a non-viable pregnancy with pooled sensitivity of 66.5% (53.6% to 77.4%), specificity of 96.3% (91.1% to 98.5%), positive likelihood ratio of 18 (7.2 to 45), and negative likelihood ratio of 0.35 (0.24 to 0.50). The probability of a non-viable pregnancy was raised from 62.9% to 96.8%.

Conclusion A single progesterone measurement for women in early pregnancy presenting with bleeding or pain and inconclusive ultrasound assessments can rule out a viable pregnancy.

Introduction

Vaginal bleeding and abdominal pain are the most common causes of consultation in early pregnancy; 30% of women will experience pain or bleeding in their first trimester.[1] These symptoms lead to anxiety and can be the first signs of a possible miscarriage or an ectopic pregnancy. Miscarriage occurs in 10-20% of all recognised pregnancies, and the prevalence of ectopic pregnancies in the United Kingdom is 1.6%, which is raised to 3% in women with symptoms.[2-4] Most women seeking medical advice have a transvaginal ultrasound scan to confirm a viable pregnancy, a miscarriage, or an ectopic pregnancy. However, even with expert use of transvaginal ultrasound, confirming if a pregnancy is intrauterine or extrauterine may not be possible in 8-31% of cases at the first visit.[5] An observational study of pregnancies with inconclusive ultrasound results has shown that 50% spontaneously resolve (decreasing

concentrations of β human chorionic gonadotrophin (β-hCG)), 27% are subsequently diagnosed as viable, and 14% are diagnosed as ectopic pregnancies.[6] The high incidence of miscarriages and ectopic pregnancies in women with inconclusive ultrasound results warrants further tests to reach a diagnosis. Measurement of serum β-hCG can be useful, but often more than one β-hCG measurements is needed to make a diagnosis.[7]

Serum progesterone has been proposed as a useful test to distinguish a viable pregnancy from a miscarriage or ectopic pregnancy.[8] Low progesterone values are associated with miscarriages and ectopic pregnancies, both considered non-viable pregnancies, and high progesterone concentrations with viable pregnancies.[8] Knowledge about the diagnostic accuracy of a single serum progesterone measurement to determine the viability of the pregnancy may accelerate diagnosis, avoiding further tests and unnecessary interventions, and improve outcomes. National recommendations refrain from defining how this test should be used in diagnostic algorithms.[8] The reason for this is that clinical practice is informed by individual small studies with unstable inferences and conflicting results on the accuracy of a single serum progesterone measurement in determining the viability of the pregnancy.[5 6 8] A meta-analysis in 1998 did not take into account the recent widespread use of transvaginal ultrasound, which may improve the accuracy of this test.[9] The aim of our review was to evaluate the accuracy of a single progesterone measurement, by a meta-analysis of diagnostic accuracy studies, to predict pregnancy outcome in women with pain or bleeding in early pregnancy with or without an inconclusive ultrasound diagnosis.

Methods

Literature search

We did a comprehensive literature search to identify studies with a population of women with spontaneous pregnancy of less than 14 weeks of gestation, in which the single serum progesterone measurement was used to predict the outcome of pregnancy (viable intrauterine pregnancy, miscarriage, or ectopic pregnancy). The databases searched included Medline, Embase, CINAHL, and the Cochrane Library from inception to April 2012. We searched Medline, Embase and CINAHL by a combination of the keywords for the test (progestor*) and outcomes of interest ("ectopic pregnancy", "tubal pregnancy", "viab* pregnancy", "failing pregnancy", "miscarr*", and "abort*") with their associated medical subject and Emtree headings. We searched the Cochrane Library with the keywords "progesterone" and "pregnancy". We placed no filters or language restrictions to ensure maximal sensitivity of the searches. We also checked the reference sections of all selected articles for relevant papers.

Study selection

We selected the studies in two rounds: firstly, on title and abstract, independently by two reviewers (IDG and MA); secondly, on full text, also independently by two reviewers (JV and NM), against pre-specified criteria. We selected studies that assessed diagnostic accuracy or derived prediction rules. We excluded case-control studies, narrative reviews, letters, editorials, comments, and case series. We used systematic reviews and meta-analyses only as a source of references. Studies needed to include women with spontaneous early pregnancy to be eligible for the review. We excluded studies including women who had conceived after treatment to induce ovulation, who had had progesterone supplementation or in

vitro fertilisation, as well as studies that included women above a gestational age of 14 weeks. The index test was a single measurement of serum progesterone. We excluded studies that did not report a cut-off value for progesterone but reported results only by high or low progesterone.

Data extraction and quality assessment

Two reviewers (JV and NM) extracted data, and a third reviewer (IDG) checked the data. From the relevant articles, we extracted information on study design, inclusion and exclusion criteria, population (asymptomatic women/women with symptoms, women with inconclusive transvaginal ultrasound results), details of progesterone test (assay, cut-off value), details of other tests (β-hCG, transvaginal ultrasound), details about pregnancy outcomes (description, diagnostic methods), flow of patients, and two by two tables for pregnancy outcomes according to progesterone test. One reviewer (JV) assessed included studies for methodological quality, and two reviewers (MA and NM) checked them. We assessed the quality of studies by using an outlined component approach for diagnostic accuracy studies (QUADAS).[10] The reference standard for our included studies was the pregnancy outcome as defined during follow-up in the original studies. For this reason, we omitted one item on the time period from index test to reference standard, as the reference standard diagnosis is normally reached within hours or days virtually eliminating possible delayed verification bias from all studies. We grouped studies according to whether women presented with pain or bleeding and inconclusive ultrasound examination or with symptoms alone.

Data synthesis and analysis

We plotted individual studies' estimates of sensitivities and specificities on summary receiver operating characteristics space and forest plots for visual examination of heterogeneity. We used the SAS statistical package to meta-analyse a pair of sensitivity and specificity from each included study by using the hierarchical summary receiver operating characteristics approach.[11 12] This approach estimates the position and shape of the summary receiver operating characteristics curve and takes into account both within and between study variations. Many studies reported progesterone concentrations in ng/mL, so we converted those in nmol/L to ng/mL. To generate a summary receiver operating characteristics curve using all available studies for each meta-analysis, where more than one cut-off value was reported in a study we chose the two by two table for the lowest cut-off value because lower cut-offs were more widely reported in included studies. When all the parameters of the hierarchical summary receiver operating characteristics model could not be estimated owing to a limited number of studies, we simplified it by assuming a symmetrical shape for the curve. We considered methods that allow joint synthesis of sensitivities and specificities at multiple thresholds, but none was appropriate because most studies reported sensitivity and specificity at a single threshold.[13-15] For meta-analysis of studies that used the same or similar cut-off values, we used parameter estimates from the models to derive summary operating points (that is, summary sensitivities and specificities), with 95% confidence regions, and summary likelihood ratios. We calculated post-test probabilities by using the summary likelihood ratios and the median prevalence values with their ranges as the pre-test probabilities.

Results

Literature identification and study quality

Figure 1⇓ shows the literature search and study selection process. We identified 4338 citations; we obtained 203 articles for full text evaluation and identified another five studies through manual checking of reference lists. We contacted 20 authors of 25 different studies for additional data. Five authors replied to our emails, but only one of these authors could provide us with additional information. Overall, 26 studies were eligible for inclusion, with a total of 9436 participants. Nineteen studies included women with pain or bleeding alone,[16-33] and seven studies included women with pain or bleeding and inconclusive ultrasound diagnosis.[34-40] The supplementary (web extra) table shows the characteristics of the included studies.

Figure 2⇓ shows the results of the quality assessment. Most of the studies of women with pain or bleeding with inconclusive ultrasound examination were of high quality; all seven studies included participants representative of the population of interest and contained adequate details about the index and reference standard tests. For pregnancies that resolved spontaneously, two studies did not investigate further to differentiate between a miscarriage and an ectopic pregnancy.[34 35] None of these seven studies adopted blinding for the reference standard results. The studies including women with pain or bleeding alone (without ultrasound scan) were of intermediate quality, and potential biases were difficult to exclude because of poor reporting of the primary studies. Progesterone assays used in the primary studies were found to be comparable and all were included. The insufficient reporting along with the limited number of studies prevented us from evaluating the effect of potential sources of heterogeneity.

Progesterone test in women with pain or bleeding and inconclusive ultrasound

Seven prospective cohort studies, including 2379 women, in women with pain or bleeding and inconclusive ultrasound diagnosis evaluated the diagnostic accuracy of the single serum progesterone measurement to predict the possibility of a viable pregnancy, miscarriage, or ectopic pregnancy. The thresholds of progesterone used ranged from 3.2 to 11 ng/mL (10 to 35 nmol/L); the most commonly used threshold was 5 ng/mL (16 nmol/L), which was used in three studies.[34 37 38] Figures 3⇓ and 4⇓ show the estimated sensitivities and specificities when using the single progesterone measurement test for differentiating between viable and non-viable pregnancies, which includes both miscarriages and ectopic pregnancies. After meta-analysis of five studies (1998 participants) with similar cut-off values (3.2 to 6 ng/mL),[34-38] we found that a single progesterone measurement predicted a non-viable pregnancy with pooled sensitivity of 74.6% (95% confidence interval 50.6% to 89.4%), specificity of 98.4% (90.9% to 99.7%), positive likelihood ratio of 45 (7.1 to 289), and negative likelihood ratio of 0.26 (0.12 to 0.57). In the studies included in this meta-analysis, the median prevalence of a non-viable pregnancy was 73.2%. However, if progesterone was lower than the cut-off value (3.2 to 6 ng/mL), the probability of a non-viable pregnancy was 99.2% compared with 44.8% if progesterone was higher. The progesterone test had a very poor predictive accuracy for diagnosing ectopic pregnancy, for which a low progesterone concentration did not rule in or out an ectopic pregnancy (table⇓).

Progesterone test in women with pain or bleeding alone

Nineteen cohort studies, including 7057 women, evaluated the diagnostic accuracy of the single serum progesterone measurement to predict pregnancy outcomes in women with pain or bleeding alone. The table⇓ shows the pooled diagnostic accuracy estimates for identifying women with non-viable pregnancies with the most commonly reported thresholds, and figure 5⇓ shows the estimates separately for each study according to available cut-off values. For women with symptoms alone, the progesterone test had a higher specificity using a threshold of 10 ng/mL (nine studies with 4689 participants[18 20 21 24 26 29-32]), rather than higher thresholds at 15 and 20 ng/mL, and predicted a non-viable pregnancy with pooled sensitivity of 66.5% (53.6% to 77.4%), specificity of 96.3% (91.1% to 98.5%), positive likelihood ratio of 18 (7.2 to 45), and negative likelihood ratio of 0.35 (0.24 to 0.50). The median prevalence of a non-viable pregnancy in the studies included in this analysis was 62.9%; this was raised to 96.8% if the progesterone was lower than 10 ng/mL compared with a 37.2% probability if the progesterone was higher.

Figure 6⇓ shows summary receiver operating characteristics curves for commonly reported cut-off values and a curve of all studies at a selected threshold for each study. This shows that for higher thresholds the specificity is lower with generally higher sensitivity. Specifically, with a threshold of 15 ng/mL (nine studies with 5128 participants[20 22-25 27 28 31 32]) this test predicted a non-viable pregnancy with pooled sensitivity of 83.3% (66.6% to 92.6%), specificity of 87.5% (78.5% to 93.1%), positive likelihood ratio of 6.7 (3.8 to 12), and negative likelihood ratio of 0.35 (0.09 to 0.5); with a threshold of 20 ng/mL (eight studies with 4348 participants[17 20 21 23 24 26 29 31]) it predicted a non-viable pregnancy with pooled sensitivity of 85.7% (72.3% to 93.2%), specificity of 66.6% (47% to 91.8%), positive likelihood ratio of 2.6 (1.5 to 4.5), and negative likelihood ratio of 0.22 (0.1 to 0.47).

In the McCord 1996 study,[24] 2248 of 3674 participants had non-viable pregnancies. This constitutes 78% of the number of participants with non-viable pregnancies in the analyses at 10 and 15 ng/mL cut-off values and 84% of the non-viable pregnancies in the analysis at 20 ng/mL. In a sensitivity analysis, the effect of removal of the study on the summary estimates at each cut-off value was negligible.

Discussion

The meta-analysis shows that a single progesterone measurement is useful in predicting non-viable pregnancies in women with pain or bleeding when an ultrasound investigation proves to be inconclusive. A low concentration of progesterone (less than 3.2 to 6 ng/mL) in these women ruled out a viable pregnancy in 99.2% of women. However, the test cannot distinguish women with an ectopic pregnancy from those with an early normal pregnancy or a miscarriage and should not be used for this purpose. For women with symptoms but without an ultrasound investigation, it may also be a useful test but is less accurate for ruling out women with a normal viable pregnancy.

Strengths and weaknesses of study

The strengths of our study are that we did an extensive systematic search of electronic databases without language restrictions, which would have captured all existing good quality studies. The high number of included studies in our meta-analyses strengthens the power of our conclusions and enabled us to explore the diagnostic accuracy of the progesterone

test for multiple cut-off values. The methodological quality of the studies included in the review was satisfactory, but some limitations of doing pragmatic studies in this clinical setting were highlighted. The classification of women according to their pregnancy outcome, which was used as the reference standard, is a problem for miscarriages and ectopic pregnancies as in many cases a firm diagnosis cannot be made. Even though the studies did not adopt a blinded assessment of the reference standard, which in our case was the final diagnosis, this is unlikely to have introduced bias when it comes to distinguishing between viable and non-viable pregnancies, as viable pregnancies lead to a live birth and most of the studies had long enough follow-up to verify this outcome. The primary studies did not adjust for known confounding factors, such as the gestational age, but this could be explained by the high prevalence of non-viable pregnancies, in which changes in progesterone concentrations are poorly described compared with viable pregnancies.[41] Assessment of quality was also hampered by unclear reporting in some studies. Poor reporting occurred in most of the studies in the description of the reference standard, reporting of uninterpretable results, and explanation of withdrawals from the study.

Many studies, especially for women with pain or bleeding who did not have an ultrasound assessment, had very different prevalences of pregnancy outcomes compared with more recent studies; this reflects the different clinical settings and populations that this test evaluated. Despite the heterogeneity, all studies reported good predictive ability of the progesterone test to differentiate viable pregnancies from miscarriages and ectopic pregnancies, especially in women with pain or bleeding with inconclusive ultrasound results, in whom the chance of miscarriages or ectopic pregnancies is high. The test, therefore, has generalisability and may be applicable in clinical practice in a variety of settings.

Strengths and weaknesses in relation to other studies

The Royal College of Obstetricians and Gynaecologists' guideline for early pregnancy suggests that the single serum progesterone measurement is a useful test to predict pregnancy outcome.[8] It states that a threshold for serum progesterone of 20 nmol/L (about 6.2 ng/mL) has a sensitivity of 93% and specificity of 94% to predict a non-viable pregnancy in women with symptoms and inconclusive ultrasound but acknowledges that no discriminatory value for this test exists to confidently rule in or out a normal viable pregnancy. This conflicts with the findings in this meta-analysis, which is based also on the studies cited by this guideline.[24 34 39] The existing evidence shows that specificity is higher (98.4%) and sensitivity lower (74.6%) using a cut-off value around 5 ng/mL. The findings of our study support those of a meta-analysis by Mol et al in 1998,[9] which focused on a single progesterone measurement to predict ectopic pregnancy. The results of that study showed that a single progesterone measurement was useful for predicting the viability of a pregnancy but not for discriminating between an ectopic and a non-ectopic pregnancy, and this agrees with our results. The clinically most relevant question of accuracy of progesterone in women with inconclusive ultrasound investigations, however, is not covered by Mol et al.

Meaning of study: implications for clinicians

The outcome of a pregnancy in women with pain or bleeding in early pregnancy cannot be determined clinically alone after inconclusive ultrasound assessments. In such situations, a chance exists of a normal viable pregnancy that may be too early to be

detected on ultrasound; our review estimates the median prevalence of this to be 26.8%. Most women, however, have a non-viable pregnancy in the form of a miscarriage or an ectopic pregnancy. Serial serum β-hCG measurements are needed to differentiate between these outcomes in most diagnostic algorithms.[7] Suboptimal rise in β-hCG (<66%) after 48 hours indicates that miscarriage or ectopic pregnancy is a possibility. However, a rise in suboptimal β-hCG may occur even in viable pregnancies in up to 19% of cases.[7] Therefore, the β-hCG test should be complemented by another test to increase its diagnostic accuracy. Serum progesterone measurement is a non-invasive test that can be done at the same time that blood is drawn for β-hCG measurements. It is already in use in many early pregnancy assessment units, although the accuracy of the progesterone test and the interpretation of the measured concentrations are uncertain. This test could be added to the existing algorithms for evaluation of early pregnancy, and its effect should be evaluated through a randomised trial comparing algorithms with and without serum progesterone.[42] Our review suggests that a single serum progesterone measurement is more accurate when it follows an inconclusive ultrasound assessment; in women with pain or bleeding who did not have an ultrasound scan, the progesterone test was less accurate in predicting viability of a pregnancy.

In conclusion, this meta-analysis found that a single progesterone measurement for women in early pregnancy presenting with bleeding or pain and inconclusive ultrasound assessments can rule out a viable pregnancy.

1   Everett C. Incidence and outcome of bleeding before the 20th week of pregnancy: prospective study from general practice. *BMJ* 1997;315:32-4.
2   Alberman E. Spontaneous abortion: epidemiology. In: Stabile S, Grudzinskas JG, Chard T, eds. Spontaneous abortion: diagnosis and treatment. Springer-Verlag, 1992:9-20.
3   Hospital Episode Statistics. NHS Maternity Statistics, England 2009-10. Figure 1: Total deliveries, miscarriages and ectopic pregnancies per 100 deliveries, 1997-98 to 2009-10 (available at www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=1804).
4   Shillito J, Walker JJ. Early pregnancy assessment units. *Br J Hosp Med* 1997;58:505-9.
5   Condous G, Okaro E, Bourne T. The conservative management of early pregnancy complications: a review of the literature. *Ultrasound Obstet Gynecol* 2003;22:420-30.
6   Banerjee S, Aslam N, Zosmer N, Woelfer B, Jurkovic D. The expectant management of women with early pregnancy of unknown location. *Ultrasound Obstet Gynecol* 1999;14:231-6.
7   Kadar N, Caldwell BV, Romero R. A method of screening for ectopic pregnancy and its indications. *Obstet Gynecol* 1981;58:162-6.
8   Hinshaw K, Fayyad A, Munjuluri P. The management of early pregnancy loss. Royal College of Obstetricians and Gynaecologists, 2006. (Green-top guideline No 25.)
9   Mol BW, Lijmer JG, Ankum WM, van der Veen F, Bossuyt PM. The accuracy of single serum progesterone measurement in the diagnosis of ectopic pregnancy: a meta-analysis. *Hum Reprod* 1998;13:3220-7.
10   Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.

What is already known on this topic

Vaginal bleeding or pain occurs in 30% of women in early pregnancy; investigations are aimed at differentiating a viable intrauterine pregnancy from a non-viable one

A single progesterone measurement in early pregnancy has been suggested to be a useful test for discriminating between viable and non-viable pregnancies

However, clinical practice is informed by individual small studies with conflicting results on the test's accuracy in determining the viability of a pregnancy

What this study adds

A single progesterone measurement can discriminate between a viable and a non-viable pregnancy

A single low progesterone measurement for women in early pregnancy presenting with bleeding or pain and inconclusive ultrasound results can rule out a viable pregnancy

11 Macaskill P. Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. *J Clin Epidemiol* 2004;57:925-32.

12 Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001;20:2865-84.

13 Hamza TH, Arends LR, van Houwelingen HC, Stijnen T. Multivariate random effects meta-analysis of diagnostic tests with multiple thresholds. *BMC Med Res Methodol* 2009;10:73.

14 Kester AD, Buntinx F. Meta-analysis of ROC curves. *Med Decis Making* 2000;20:430-9.

15 Dukic V, Gatsonis C. Meta-analysis of diagnostic test accuracy assessment studies with varying number of thresholds. *Biometrics* 2003;59:936-46.

16 Buckley RG, King KJ, Disney JD, Riffenburgh RH, Gorman JD, Klausen JH. Serum progesterone testing to predict ectopic pregnancy in symptomatic first-trimester patients. *Ann Emerg Med* 2000;36:95-100.

17 Camejo MI, Proverbio F, Febres F, Casart YC. Bioactive to immunoreactive ratio of circulating human chorionic gonadotropin as possible evaluation for the prognosis of threatened abortion. *Eur J Obstet Gynecol Reprod Biol* 2003;109:181-4.

18 Choe JK, Check JH, Nowroozi K, Benveniste R, Barnea ER. Serum progesterone and 17-hydroxyprogesterone in the diagnosis of ectopic pregnancies and the value of progesterone replacement in intrauterine pregnancies when serum progesterone levels are low. *Gynecol Obstet Invest* 1992;34:133-8.

19 Dart R, Dart L, Segal M, Page C, Brancato J. The ability of a single serum progesterone value to identify abnormal pregnancies in patients with beta-human chorionic gonadotropin values less than 1,000 mIU/mL. *Acad Emerg Med* 1998;5:304-9.

20 Gelder MS, Boots LR, Younger JB. Use of a single random serum progesterone value as a diagnostic aid for ectopic pregnancy. *Fertil Steril* 1991;55:497-500.

21 Grosskinsky CM, Hage ML, Tyrey L, Christakos AC, Hughes CL. hCG, progesterone, alpha-fetoprotein, and estradiol in the identification of ectopic pregnancy. *Obstet Gynecol* 1993;81:705-9.

22 Hubinont CJ, Thomas C, Schwers JF. Luteal function in ectopic pregnancy. *Am J Obstet Gynecol* 1987;156:669-74.

23 Kratzer PG, Taylor RN. Corpus luteum function in early pregnancies is primarily determined by the rate of change of human chorionic gonadotropin levels. *Am J Obstet Gynecol* 1990;163:1497-502.

24 McCord ML, Muram D, Buster JE, Arheart KL, Stovall TG, Carson SA. Single serum progesterone as a screen for ectopic pregnancy: exchanging specificity and sensitivity to obtain optimal test performance. *Fertil Steril* 1996;66:513-6.

25 Mesrogli M, Degenhardt F, Maas DH, Klaus I, Busche M, Schneider J. [Tubal pregnancies: early pregnancy factor, progesterone, beta-HCG and vaginal sonography as differential diagnostic parameters] [German]. *Z Geburtshilfe Perinatol* 1988;192:130-2.

26 O'Leary P, Nichols C, Feddema P, Lam T, Aitken M. Serum progesterone and human chorionic gonadotrophin measurements in the evaluation of ectopic pregnancy. *Aust N Z J Obstet Gynaecol* 1996;36:319-23.

27 Perkins SL, Al-Ramahi M, Claman P. Comparison of serum progesterone as an indicator of pregnancy nonviability in spontaneously pregnant emergency room and infertility clinic patient populations. *Fertil Steril* 2000;73:499-504.

28 Peterson CM, Kreger D, Delgado P, Hung TT. Laboratory and clinical comparison of a rapid versus a classic progesterone radioimmunoassay for use in determining abnormal and ectopic pregnancies. *Am J Obstet Gynecol* 1992;166:562-6.

29 Sauer MV, Sinosich MJ, Yeko TR, Vermesh M, Buster JE, Simon JA. Predictive value of a single serum pregnancy associated plasma protein-A or progesterone in the diagnosis of abnormal pregnancy. *Hum Reprod* 1989;4:331-4.

30 Takahashi K, Yamane Y, Kitao M. Prognostic value of serum CA 125 and serum progesterone in threatened abortion. *J Obstet Gynaecol* 1992;12:147-9.

31 Williams RS, Gaines IL, Fossum GT. Progesterone in diagnosis of ectopic pregnancy. *J Fla Med Assoc* 1992;79:237-9.

32 Witt BR, Wolf GC, Wainwright CJ, Johnston PD, Thorneycroft IH. Relaxin, CA-125, progesterone, estradiol, Schwangerschaft protein, and human chorionic gonadotropin as predictors of outcome in threatened and nonthreatened pregnancies. *Fertil Steril* 1990;53:1029-36.

33 Yeko TR, Rodi IA, Gorrill MJ, Buster JE, Hughes LH, Sauer MV. Timely diagnosis of early ectopic pregnancy using a single blood progesterone measurement. *Fertil Steril* 1987;48:1048-50.

34 Banerjee S, Aslam N, Zosmer N, Woelfer B, Jurkovic D. The expectant management of women with early pregnancy of unknown location. *Ultrasound Obstet Gynecol* 1999;14:231-6.

35 Dart R, Ramanujam P, Dart L. Progesterone as a predictor of ectopic pregnancy when the ultrasound is indeterminate. *Am J Emerg Med* 2002;20:575-9.

36 Day A, Sawyer E, Mavrelos D, Tailor A, Helmy S, Jurkovic D. Use of serum progesterone measurements to reduce need for follow-up in women with pregnancies of unknown location. *Ultrasound Obstet Gynecol* 2009;33:704-10.

37 El Bishry G, Ganta S. The role of single serum progesterone measurement in conjunction with beta hCG in the management of suspected ectopic pregnancy. *J Obstet Gynaecol* 2008;28:413-7.

38 Florio P, Severi FM, Bocchi C, Luisi S, Mazzini M, Danero S, et al. Single serum activin a testing to predict ectopic pregnancy. *J Clin Endocrinol Metab* 2007;92:1748-53.

39 Hahlin M, Sjoblom P, Lindblom B. Combined use of progesterone and human chorionic gonadotropin determinations for differential diagnosis of very early pregnancy. *Fertil Steril* 1991;55:492-6.

40 Valley VT, Mateer JR, Aiman EJ, Thoma ME, Phelan MB. Serum progesterone and endovaginal sonography by emergency physicians in the evaluation of ectopic pregnancy. *Acad Emerg Med* 1998;5:309-13.

41 Mol BW, Hajenius PJ, Engelsbel S, Ankum WM, Hemrika DJ, Bossuyt PMM, et al. Are gestational age and endometrial thickness alternatives for serum hCG measurement in the diagnosis of ectopic pregnancy? *Fertil Steril* . 1999;72:643-5.

42 Bossuyt PM, Lijmer JG, Mol BW. Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet* 2000;356:1844-7.

Table

Table 1| Summary estimates for each pregnancy outcome at different thresholds in women with pain and/or bleeding with inconclusive ultrasound diagnosis and for women with pain and/or bleeding alone

Studies (No of participants)	Cut-off (ng/mL)	Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)	Pre-test and post-test probabilities (ranges)*		
						Before test (%)	After test if positive (%)	After test if negative (%)
Women with pain and/or bleeding with inconclusive ultrasound diagnosis								
Progesterone test to predict ectopic *v* non-ectopic pregnancy:								
5 studies (1998 women)	<3.2-6	36.5 (12 to 70.6)	41.7 (16 to 72.9)	0.63 (0.29 to 1.35)	1.52 (0.99 to 2.36)	14.2 (7.7-28.6)	9.44 (5-20.2)	20.1 (11.3-37.8)
Progesterone test to predict non-viable *v* viable pregnancy:								
5 studies (1998 women)	<3.2-6	74.6 (50.6 to 89.4)	98.4 (90.9 to 99.7)	45.4 (7.13 to 288.9)	0.26 (0.12 to 0.57)	73.2 (71.1-85.9)	99.2 (99.1-99.6)	44.8 (39-61.3)
Women with pain and/or bleeding alone								
Progesterone test to predict non-viable *v* viable pregnancy:								
9 studies (4689 women)	<10	66.5 (53.6 to 77.4)	96.3 (91.1 to 98.5)	18 (7.19 to 44.8)	0.35 (0.24 to 0.50)	62.9 (47.4-71.8)	96.8 (94.2-97.9)	37.2 (24-47.1)
9 studies (5128 women)	<15	83.3 (66.6 to 92.6)	87.5 (78.5 to 93.1)	6.66 (3.81 to 11.7)	0.19 (0.09 to 0.40)	70 (12.6-79.7)	93.9 (49-96.3)	30.7 (2.7-42.7)
8 studies (4348 women)	<20	85.7 (72.3 to 93.2)	66.6 (47.0 to 81.8)	2.56 (1.46 to 4.50)	0.22 (0.10 to 0.47)	61.2 (30.8-71.8)	80.2 (53.3-86.7)	25.8 (8.9-35.9)

*Pre-test probabilities are median prevalence values with corresponding ranges.

Figures
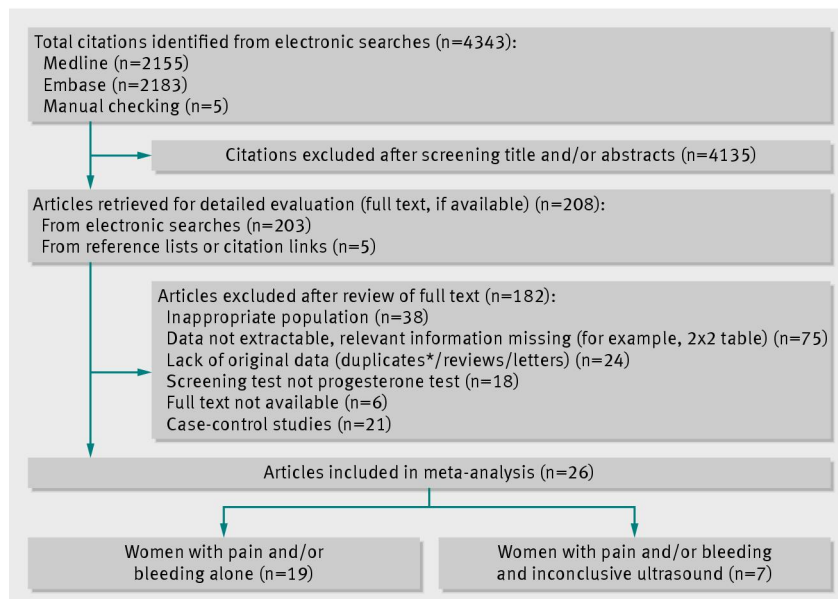
Fig 1 Flow chart of study selection. *Three studies by Stovall et al reported same patients as study by McCord et al. Study by McCord was included, because it reported on more participants
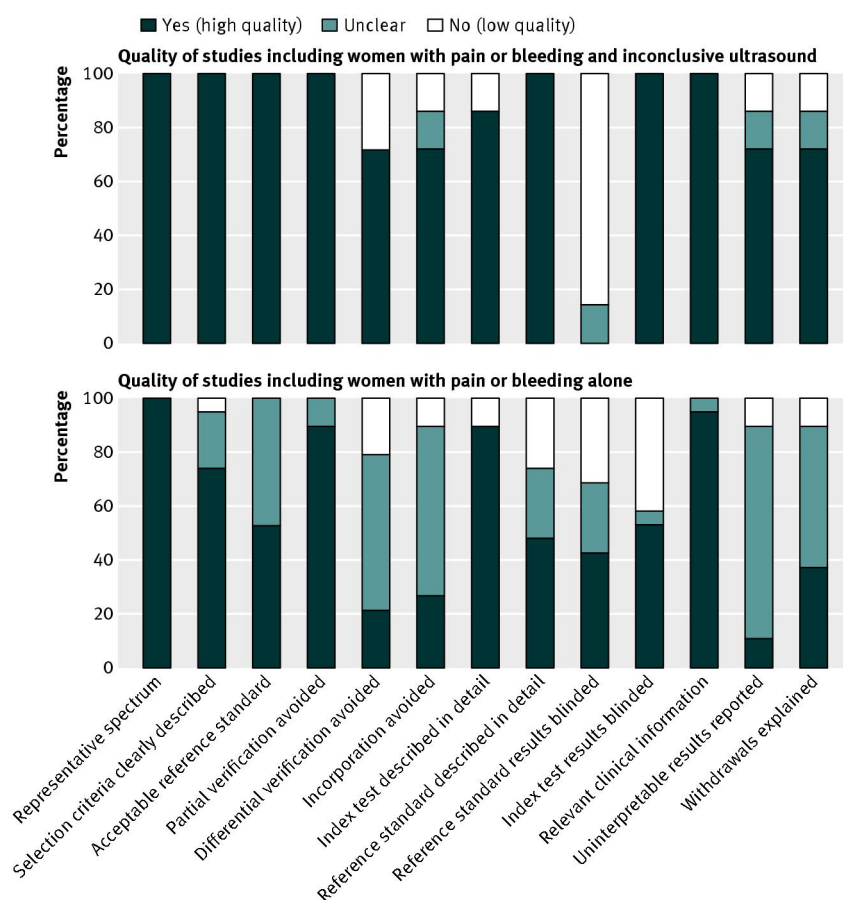
Fig 2 Cumulative bar plot of methodological quality items across studies including women with pain or bleeding and inconclusive ultrasound (top) and women with pain or bleeding alone (bottom)

Study	TP	FP	FN	TN	Cut-off (ng/mL)	Sensitivity	Specificity	Sensitivity	Specificity
Progesterone test to predict a non-viable *v* viable pregnancy									
Banerjee 1999	51	0	42	34	3.2			0.55 (0.44 to 0.65)	1.00 (0.90 to 1.00)
Dart 2002	77	0	8	14	5			0.91 (0.82 to 0.96)	1.00 (0.77 to 1.00)
El Bishry 2008	51	1	40	34	5			0.56 (0.45 to 0.66)	0.97 (0.85 to 1.00)
Florio 2007	316	1	65	154	5			0.83 (0.79 to 0.87)	0.99 (0.96 to 1.00)
Day 2009	656	9	206	239	6.3			0.76 (0.73 to 0.79)	0.96 (0.93 to 0.98)
Hahlin 1991	180	1	54	72	9.4			0.77 (0.71 to 0.82)	0.99 (0.93 to 1.00)
Valley 1998	49	1	8	10	11			0.86 (0.74 to 0.94)	0.91 (0.59 to 1.00)
Progesterone test to predict ectopic *v* non-ectopic pregnancy									
Banerjee 1999	2	49	16	60	3.2			0.11 (0.01 to 0.35)	0.55 (0.45 to 0.65)
Dart 2002	19	58	2	20	5			0.90 (0.70 to 0.99)	0.26 (0.16 to 0.37)
El Bishry 2008	16	36	20	54	5			0.44 (0.28 to 0.62)	0.60 (0.49 to 0.70)
Florio 2007	12	305	64	155	5			0.16 (0.08 to 0.26)	0.34 (0.29 to 0.38)
Day 2009	28	637	57	388	6.3			0.33 (0.23 to 0.44)	0.38 (0.35 to 0.41)
Hahlin 1991	21	29	4	14	9.4			0.84 (0.64 to 0.95)	0.33 (0.19 to 0.49)
Valley 1998	119	62	40	86	11			0.75 (0.67 to 0.81)	0.58 (0.50 to 0.66)

0 0.2 0.4 0.6 0.8 1.0   0 0.2 0.4 0.6 0.8 1.0

Fig 3 Forest plot of study results of progesterone test in women with pain or bleeding and inconclusive ultrasound assessment grouped according to outcome. FN=false negative; FP=false positive; TN=true negative; TP=true positive

Fig 4 Summary receiver operating characteristics plot of progesterone test at cut-off values between 3.2 and 6.4 ng/mL used to identify non-viable pregnancies in women with pain or bleeding and inconclusive ultrasound assessment (black dot=summary sensitivity and specificity; dotted region around it=95% confidence region)

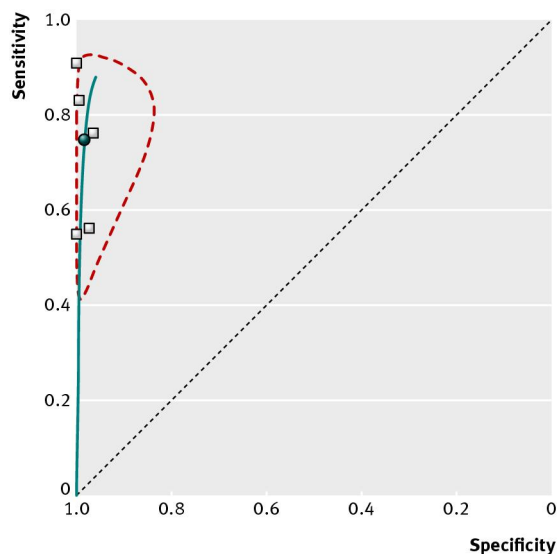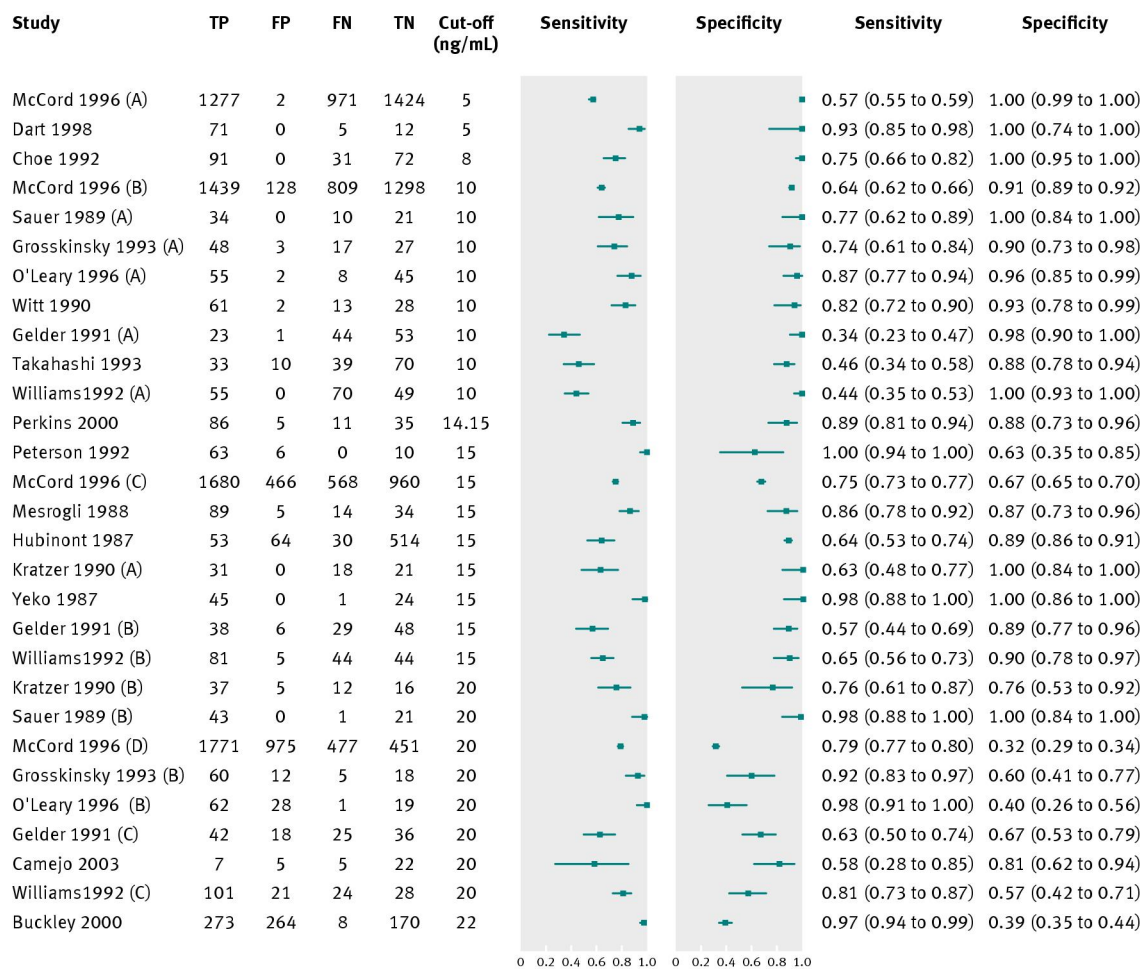| Study	TP	FP	FN	TN	Cut-off (ng/mL)	Sensitivity	Specificity	Sensitivity	Specificity
McCord 1996 (A)	1277	2	971	1424	5			0.57 (0.55 to 0.59)	1.00 (0.99 to 1.00)
Dart 1998	71	0	5	12	5			0.93 (0.85 to 0.98)	1.00 (0.74 to 1.00)
Choe 1992	91	0	31	72	8			0.75 (0.66 to 0.82)	1.00 (0.95 to 1.00)
McCord 1996 (B)	1439	128	809	1298	10			0.64 (0.62 to 0.66)	0.91 (0.89 to 0.92)
Sauer 1989 (A)	34	0	10	21	10			0.77 (0.62 to 0.89)	1.00 (0.84 to 1.00)
Grosskinsky 1993 (A)	48	3	17	27	10			0.74 (0.61 to 0.84)	0.90 (0.73 to 0.98)
O'Leary 1996 (A)	55	2	8	45	10			0.87 (0.77 to 0.94)	0.96 (0.85 to 0.99)
Witt 1990	61	2	13	28	10			0.82 (0.72 to 0.90)	0.93 (0.78 to 0.99)
Gelder 1991 (A)	23	1	44	53	10			0.34 (0.23 to 0.47)	0.98 (0.90 to 1.00)
Takahashi 1993	33	10	39	70	10			0.46 (0.34 to 0.58)	0.88 (0.78 to 0.94)
Williams1992 (A)	55	0	70	49	10			0.44 (0.35 to 0.53)	1.00 (0.93 to 1.00)
Perkins 2000	86	5	11	35	14.15			0.89 (0.81 to 0.94)	0.88 (0.73 to 0.96)
Peterson 1992	63	6	0	10	15			1.00 (0.94 to 1.00)	0.63 (0.35 to 0.85)
McCord 1996 (C)	1680	466	568	960	15			0.75 (0.73 to 0.77)	0.67 (0.65 to 0.70)
Mesrogli 1988	89	5	14	34	15			0.86 (0.78 to 0.92)	0.87 (0.73 to 0.96)
Hubinont 1987	53	64	30	514	15			0.64 (0.53 to 0.74)	0.89 (0.86 to 0.91)
Kratzer 1990 (A)	31	0	18	21	15			0.63 (0.48 to 0.77)	1.00 (0.84 to 1.00)
Yeko 1987	45	0	1	24	15			0.98 (0.88 to 1.00)	1.00 (0.86 to 1.00)
Gelder 1991 (B)	38	6	29	48	15			0.57 (0.44 to 0.69)	0.89 (0.77 to 0.96)
Williams1992 (B)	81	5	44	44	15			0.65 (0.56 to 0.73)	0.90 (0.78 to 0.97)
Kratzer 1990 (B)	37	5	12	16	20			0.76 (0.61 to 0.87)	0.76 (0.53 to 0.92)
Sauer 1989 (B)	43	0	1	21	20			0.98 (0.88 to 1.00)	1.00 (0.84 to 1.00)
McCord 1996 (D)	1771	975	477	451	20			0.79 (0.77 to 0.80)	0.32 (0.29 to 0.34)
Grosskinsky 1993 (B)	60	12	5	18	20			0.92 (0.83 to 0.97)	0.60 (0.41 to 0.77)
O'Leary 1996 (B)	62	28	1	19	20			0.98 (0.91 to 1.00)	0.40 (0.26 to 0.56)
Gelder 1991 (C)	42	18	25	36	20			0.63 (0.50 to 0.74)	0.67 (0.53 to 0.79)
Camejo 2003	7	5	5	22	20			0.58 (0.28 to 0.85)	0.81 (0.62 to 0.94)
Williams1992 (C)	101	21	24	28	20			0.81 (0.73 to 0.87)	0.57 (0.42 to 0.71)
Buckley 2000	273	264	8	170	22			0.97 (0.94 to 0.99)	0.39 (0.35 to 0.44)

0 0.2 0.4 0.6 0.8 1.0   0 0.2 0.4 0.6 0.8 1.0

Fig 5 Forest plot of study results of progesterone test at various cut-off values used to identify non-viable pregnancies in women with pain or bleeding alone. Study names with suffixes A to D reported accuracy of progesterone at more than one cut-off. FN=false negative; FP=false positive; TN=true negative; TP=true positive
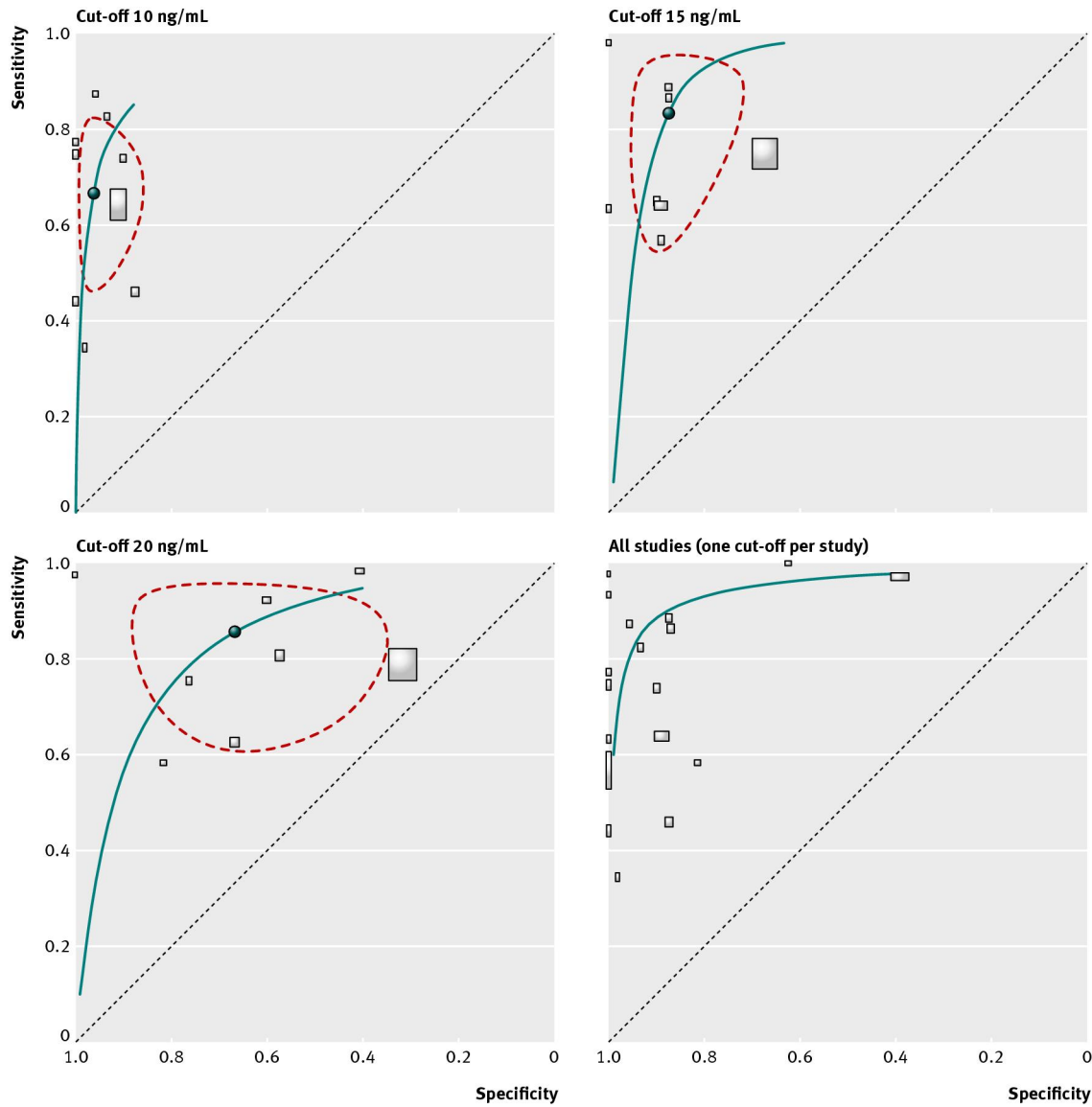
Fig 6 Summary receiver operating characteristics plot of progesterone test at different cut-off values used to identify non-viable pregnancies in women with pain or bleeding alone