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

Verhaegen J, Gallos ID, van Mello NM, Abdel-Aziz M, Takwoingi Y, Harb H, Deeks JJ, Mol BW, Coomarasamy A

CRD summary
This review concluded that a single progesterone measurement for women in early pregnancy who presented with bleeding or pain and inconclusive ultrasound assessments could rule out a viable pregnancy. This was a generally well-conducted review and the conclusion is likely to be reliable.

Authors' objectives
To determine the accuracy with which a single progesterone measurement in early pregnancy discriminated between viable and non-viable pregnancy.

Searching
MEDLINE, EMBASE, CINAHL, Web of Science, ProQuest, Conference Proceedings Citation Index and The Cochrane Library were searched, without language restrictions, from inception to April 2012; search terms were reported. Bibliographies of included articles, systematic reviews and meta-analyses were also scanned for additional studies.

Study selection
Studies that assessed diagnostic accuracy or derived prediction rules in a population of women with spontaneous pregnancy (of less than 14 weeks of gestation) were eligible for inclusion. The index test/predictive variable of interest was a single serum progesterone measurement where a specific cut-off for a positive test was reported. Case-control studies, case series, studies of women who conceived after fertility treatment or progesterone supplementation, and studies of women beyond 14 weeks gestation were excluded. The thresholds of progesterone used across the studies ranged from 3.2 to 22ng/ml. The test was used to identify non-viable and ectopic pregnancies.

Two independent reviewers selected studies for the review.

Assessment of study quality
Quality was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool. The item that assessed progression bias was omitted from the assessment.

Quality was assessed by one reviewer and checked by a second reviewer.

Data extraction
Data were extracted from each study in order to construct 2x2 tables of test performance. Sensitivity and specificity with 95% confidence intervals (CI) were calculated.

Data were extracted by one reviewer and checked by a second.

Methods of synthesis
Studies were grouped according to whether women presented with pain or bleeding and inconclusive ultrasound examination, or with symptoms alone. Hierarchical summary receiver operating characteristic (HSROC) curves were produced, and summary estimates of sensitivity and specificity with 95% confidence regions produced. Results from studies were presented on forest plots with the pooled estimate omitted to allow a visual inspection of heterogeneity; no results from statistical tests of heterogeneity were presented. Analyses were conducted for different cut-off values for the progesterone measurement. In order to include all studies in one analysis, where more than one cut-off was used, the lowest was selected. Where all parameters of the HSROC model could not be estimated, a symmetrical curve was assumed. The summary estimates of sensitivity and specificity were used to calculate summary positive and negative likelihood ratios. These, along with the median prevalence values, were used to calculate post-test probabilities.
Results of the review
Twenty-six studies met the inclusion criteria (9,436 pregnant women). Seven studies included women with pain or bleeding and an inconclusive ultrasound; most were of high quality. All seven included participants representative of the population of interest and contained adequate details about the index and reference standard tests; none blinded interpreters of the reference standard. Nineteen studies included women with pain or bleeding alone; these were considered of intermediate quality, primarily due to poor reporting making the potential for bias uncertain.

Women with pain, bleeding and an inconclusive ultrasound (2,379 participants): When predicting a non-viable pregnancy, sensitivity ranged from 55% to 91% and specificity from 91% to 100%. The pooled estimate (five studies) of sensitivity was 74.6% (95% CI 50.6 to 89.4), specificity was 98.4% (95% CI 90.9 to 99.7), Positive likelihood ratio was 45 (95% CI 7.1 to 289), and negative likelihood ratio was 0.26 (95% CI 0.12 to 0.57) with cut-off values of 3.2 to 6ng/ml. The progesterone test was a very poor predictor of ectopic pregnancy; sensitivity ranged from 11% to 90% and specificity from 26% to 60%.

Women with pain or bleeding alone (7,057 participants): When predicting a non-viable pregnancy, sensitivity ranged from 34% to 100% and specificity from 32% to 100%. When predicting a non-viable pregnancy (nine studies; 4,689 participants), pooled sensitivity was 66.5% (95% CI 53.6% to 77.4%), specificity was 96.3% (95% CI 91.1% to 98.5%), positive likelihood ratio was was 18 (95% CI 7.2 to 45), and negative likelihood ratio was 0.35 (95% CI 0.24 to 0.50) when the cut-off for a positive test was 10ng/ml. This threshold had a higher specificity than the 15 and 20ng/ml thresholds.

Results for a range of cut-offs and other sensitivity analyses were reported.

Authors’ conclusions
A single progesterone measurement for women in early pregnancy who presented with bleeding or pain and inconclusive ultrasound assessments could rule out a viable pregnancy.

CRD commentary
The authors addressed a clear research question with reproducible inclusion criteria. Several relevant sources were searched for studies in any language. It appeared that unpublished studies were not sought, so there was no specific attempt to reduce the potential for publication bias. No filters of diagnostic or prognostic terminology were used during the electronic searches, which reduced the likelihood of missing studies. Each stage of the review process was at least checked by a second reviewer, which reduced the risk of error and bias. Appropriate criteria were used to assess study quality, but results were not reported in full, so it was unclear which studies suffered from which biases. Appropriate methods of synthesis were employed with suitable subgroup and sensitivity analyses. This was a generally well-conducted review and the conclusion is likely to be reliable.

Implications of the review for practice and research
Practice: The authors stated that a single measurement of serum progesterone could be added to the algorithm for evaluation of early pregnancy, with blood being drawn for the beta-hCG measurement.

Research: The authors stated that the adding of a single measurement of serum progesterone to the algorithm for evaluation of early pregnancy should be evaluated in a randomised control trial.

Funding
None

Bibliographic details

DOI
10.1136/bmj.e6077