The accuracy of single serum progesterone measurement in the diagnosis of ectopic pregnancy: a meta-analysis


Authors' objectives
To determine the accuracy of single serum progesterone measurement in the distinction between pregnancy failure and viable intra-uterine pregnancies, and between ectopic pregnancy and all other pregnancies.

Searching
MEDLINE (from January 1966 to December 1966) and EMBASE (January 1988 to December 1996) were searched and selected studies were checked for references. The keywords for the search were 'ectopic pregnancy', 'tubal pregnancy' and 'progesterone'.

Study selection
Study designs of evaluations included in the review
No inclusion criteria relating to the study design were specified. Prospective cohort, prospective case-control studies, retrospective cohort studies and retrospective case-control studies were included in the review.

Specific interventions included in the review
Studies of single serum progesterone measurement were included. The threshold levels for both diagnostic questions ranged from 8 to 45 ng/mL.

Reference standard test against which the new test was compared
No inclusion criteria relating to the reference standard were defined. The final diagnosis was treated as the reference standard for the included studies.

Participants included in the review
Articles comparing serum progesterone concentrations in patients with ectopic pregnancy, non-viable intra-uterine pregnancy, and viable intra-uterine pregnancy were eligible for inclusion. Articles were excluded if they only compared patients with viable intra-uterine pregnancies and patients with ectopic pregnancies after the exclusion of patients with non-viable intra-uterine pregnancies.

Outcomes assessed in the review
Articles with insufficient data to construct a 2x2 contingency table of test result and final diagnosis were excluded. The outcome measures reported in the review were sensitivity and specificity.

How were decisions on the relevance of primary studies made?
The authors did not state how the papers were selected for the review, or how many reviewers performed the selection.

Assessment of study quality
Each study was scored on the following characteristics: sampling (consecutive patients versus other), data collection (prospective or retrospective) and study design (cohort versus case control). The authors did not state how the papers were assessed for validity, or how many reviewers performed the validity assessment.

Data extraction
Two authors independently constructed 2x2 tables. In the event of any disagreement, the judgement of a third author was decisive.
Methods of synthesis
How were the studies combined?
The sensitivity and specificity were calculated from the published data for each study, separately, and plotted in receiver operating characteristic (ROC) space. In cases where homogeneity could not be rejected, summary estimates of sensitivity and specificity were calculated. Where it was possible to extract data on the performance of a single serum progesterone measurement at more than one threshold from a study, data on different thresholds were used, up to a maximum of four. To control for the increased weight of these studies the number of patients was divided by the number of cut-off points. In the absence of a threshold effect (see below) summary ROC curves were constructed. To correct for bias resulting from the use of multiple 2x2 contingency tables from one study, a summary ROC curve was estimated using a random-effects regression model.

How were differences between studies investigated?
Homogeneity of the studies was tested using a chi-squared test. In case of heterogeneity, logistic regression was used to evaluate if the study characteristics, (i.e. sampling, data collection and study design) were associated with the discriminative capacity of a single serum progesterone measurement; the presence or absence of disease was used as the dependent variable. If one of the study characteristics was found to have a statistically significant impact on the performance of the test, further analysis was performed in subgroups of patients. A Spearman correlation coefficient was calculated for the association between sensitivity and specificity, to explore possible heterogeneity due to a shift in the threshold levels of serum progesterone measurement used to define a positive test. To evaluate whether the accuracy of serum progesterone measurement had changed over time, the diagnostic odds ratio (ratio of true positives to false negatives divided by the ratio of false positives to true negatives) was calculated for each study and plotted against the year of publication. A regression coefficient between the reported diagnostic odds ratio and the year of publication was calculated.

Results of the review
A total of 26 studies were included: 12 prospective cohort studies, 3 prospective case-control studies, 9 retrospective cohort studies and 2 retrospective case-control studies. Sampling of the patients was consecutive in 19 studies; other methods were used in the remaining 8 studies.

Use of single serum progesterone measurement in distinguishing between pregnancy failure and viable intra-uterine pregnancy.

The sensitivity ranged from 0.32 (specificity 0.98) to 1.0 (specificity 0.73), and the specificity from 0.52 (sensitivity 0.92) to 1.0 (sensitivity 0.35 to 0.98). A plot of the studies indicated heterogeneity, which was confirmed by a chi-squared test (P<0.001 for both sensitivity and specificity). A logistic regression analysis showed that the performance of the single serum progesterone measurement varied significantly between consecutive and non-consecutive patient series (P=0.003), between prospective and retrospective studies (P<0.001), and between cohort and case-control studies (P=0.006). A further analysis was limited to the 11 prospective cohort studies in which patients were sampled prospectively. The chi-squared test indicated heterogeneity in the sensitivity and specificity between these studies (P<0.001). The Spearman correlation coefficient was -0.64. Based on this a summary ROC curve was estimated, using only data from prospective cohort studies in which patient sampling was consecutive. This showed that single serum progesterone has almost perfect specificity with a sensitivity of 60%, and can be 95% sensitive with a specificity of 40%. The diagnostic odds ratios did not vary over time (R=0.001, P=0.97).

Use of single serum progesterone measurement in distinguishing between ectopic and non-ectopic pregnancy.

Twenty-two studies reported on this outcome. The sensitivity ranged from 0.17 (specificity 0.79) to 1.0 (specificity 0.28 to 0.60), and the specificity from 0.28 (sensitivity 1.0) to 1.0 (sensitivity 0.71). A plot of the studies indicated heterogeneity, which was confirmed by a chi-squared test (P<0.001 for both sensitivity and specificity). A logistic regression analysis showed that the performance of the single serum progesterone measurement varied significantly between the cohort and case-control studies (P<0.001), but not between prospective and retrospective (P=0.09) and consecutive and non-consecutive studies (P=0.89). A further analysis was limited to cohort studies. The chi-squared test indicated heterogeneity in the sensitivity and specificity between these studies (P<0.001). The Spearman correlation coefficient was -0.66. Based on this a summary ROC curve was estimated, using only data from cohort studies. This
showed that single serum progesterone measurement has a specificity of greater than 90% only with a sensitivity of 15%, and a sensitivity of 95% with a specificity of less than 40%. The diagnostic odds ratios did not vary over time (R=0.21, P=0.97).

**Authors’ conclusions**
A single serum progesterone measurement may serve to identify patients with first trimester pregnancy failure, including ectopic pregnancy, but it is insufficient for establishing a definite diagnosis.

**CRD commentary**
Overall, this was a good review of the area. An appropriate literature search was conducted. However, this was limited to two bibliographic databases and no attempt to identify unpublished literature was reported; relevant studies may therefore have been omitted. The inclusion criteria were clearly stated and a limited validity assessment was undertaken. The description of the review methodology was limited, although the extraction of 2x2 data (a particularly error-prone area in systematic reviews of diagnostic accuracy studies) was conducted independently, in duplicate. Details of the included studies were clearly reported and the statistical analyses were appropriate, rigorous and clear. It was sensible to limit analysis to cohort studies, as cohort studies are methodologically more likely to provide accurate estimates of the true sensitivity and specificity.

**Implications of the review for practice and research**
Practice: Patients with first trimester pregnancy failure may be identified through a single measurement of serum progesterone level. However, this study does not provide an estimate of the threshold level that provides the most acceptable levels of sensitivity and specificity. Variation in patient populations is likely to generate differences in optimal threshold levels, thus each centre should validate its own threshold values for serum progesterone levels.

Research: The authors did not state any implications for further research.

**Bibliographic details**

**PubMedID**
9853884

**Original Paper URL**
http://humrep.oxfordjournals.org/cgi/reprint/13/11/3220

**Indexing Status**
Subject indexing assigned by NLM

**MeSH**
Female; Humans; Meta-Analysis as Topic; Pregnancy; Pregnancy, Ectopic /blood /diagnosis; Progesterone /blood; ROC Curve; Sensitivity and Specificity

**AccessionNumber**
11998002121

**Date bibliographic record published**
30/04/2004

**Date abstract record published**
30/04/2004
Record Status
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.