Ultrasound tests of ovarian reserve: a systematic review of accuracy in predicting fertility outcomes
Gibreel A, Maheshwari A, Bhattacharya S, Johnson NP

CRD summary
The review assessed ultrasound-based tests for predicting fertility outcomes. It found that antral follicle count (cut-off value below four follicles) had high specificity for predicting cycle cancellation in assisted reproduction, whilst ovarian volume (cut-off value of 3 ml³) had high specificity for predicting non-pregnancy and cycle cancellation. Despite some limitations in the review process, the authors' conclusions reflect the data presented.

Authors' objectives
To assess the diagnostic accuracy of ultrasound based tests of ovarian reserve in predicting fertility outcomes.

Searching
PubMed (1980 to 2007), EMBASE (1980 to 2007), the Cochrane Library (2007:4), EBM Review (1980 to 2007) and CAB Abstracts (1990 to 2007) were searched. Search terms were reported. Reference lists of primary studies and review articles were screened for additional studies. There were no language restrictions.

Study selection
Studies that assessed the accuracy of antral follicle count, ovarian volume and stromal blood flow (or any other ultrasound test of ovarian reserve) for predicting fertility outcomes, where fertility outcomes were the reference standard, were eligible for inclusion.

Studies of inter-cycle variability or different ultrasound techniques were excluded, unless data for fertility predictors were reported for the cycles/women. Studies comparing ranges of test results between groups of women (e.g. pregnant and non-pregnant; those with good and poor response) without determining the accuracy of the test at a specified cut-off, were also excluded.

The majority of included studies were of women undergoing in vitro fertilisation (IVF). Diagnostic cut-off values and timing of the test varied across studies of the same test. The reference standard also varied across studies and included cycle cancellation, poor response or pregnancy.

Studies were assessed for inclusion by two reviewers independently; disagreements were resolved by discussion with a third/fourth reviewer.

Assessment of study quality
The authors stated that they extracted data on study quality, but no further details were reported.

Data extraction
Data were extracted to construct 2x2 contingency tables of diagnostic performance (numbers of true positive, false negative, false positive and false negative test results). Tests were classified positive if the result was abnormal, indicating reduced ovarian reserve. Disease positive (reference standard) was defined as absence of the desired fertility outcome. Corresponding authors were contacted where insufficient data were reported to populate 2x2 tables.

Data extraction was performed independently by two reviewers.

Methods of synthesis
Pooled estimates sensitivity, specificity, positive and negative likelihood ratio, and diagnostic odds ratio and their 95% confidence intervals (CIs) were calculated, where data from at least two studies using same test and reference standard, on a similar population, were available. Where studies reported several cut-off values, the lowest was used in meta-
analysis to avoid duplication. The model used to generate pooled estimates was not reported.

Statistical heterogeneity was assessed using the $\chi^2$ test. Meta-analyses were performed separately for subgroups of studies with the same cut-off value. Summary receiver operating characteristic (sROC) curves for different ultrasound tests were plotted when a Spearman correlation coefficient (between sensitivity and 1-specificity) appeared to be statistically significant, i.e. when threshold effect was deemed to be present (performance of the test varied with diagnostic cut-off).

**Results of the review**

**Antral follicle count**: Seventeen studies, with a total of 2,801 women, assessed antral follicle count; 14 studies reported data to populate 2x2 tables. Results were reported in tables for antral follicle count cut-off values ranging from four to 11 follicles, although the text focused only on the results for cut-off values of four; the extremes of the range of possibilities are illustrated below.

At an antral follicle count cut-off value of four, non-pregnancy prediction (two studies; 521 cycles) sensitivity was 12% (95% CI 9 to 16), while specificity was 98% (95% CI 95 to 99). For cycle cancellation prediction at a cut-off of four (two studies; 521 cycles), sensitivity was 67% (95% CI 48 to 82), while specificity was 95% (95% CI 92 to 96).

At an antral follicle count cut-off of 11, non-pregnancy prediction (four studies; 894 cycles) sensitivity was 74% (95% CI 70 to 77), but specificity was 34% (95% CI 29 to 40). For cycle cancellation prediction at a cut-off of 11 (two studies; 483 cycles), sensitivity was 10% (95% CI 1 to 31), but specificity was 34% (95% CI 30 to 38).

Women with an antral follicle count of less than four were 8.7 times (95% CI 2.4 to 31.7) more likely not to be pregnant, and 37 times (95% CI 13.68 to 100.45) more likely to have their cycle cancelled, than women with an antral follicle count of four or more.

Data for likelihood ratios and diagnostic odds ratios were also reported.

**Ovarian volume**: Six studies, with a total of 766 women, assessed ovarian volume. At a cut-off ovarian volume value of 3cm$^3$, the summary estimate of specificity for predication of cycle cancellation was 92% (95% CI 89 to 94) and the summary estimate of specificity for non-pregnancy was 93% (95% CI 87 to 97), based on four studies, with a total of 626 women. No other meta-analytic pooling was deemed possible.

**Ovarian stromal blood flow**: Six studies, with a total of 440 women, assessed stromal blood flow. Three studies used the same index and a similar cut-off value (presence or absence of stromal blood flow) to predict the reference standard of non-pregnancy, but meta-analysis was not undertaken due to the presence of statistical heterogeneity. Results of individual studies were not reported.

**Authors’ conclusions**

Antral follicle count, at a cut-off value of less than four, had high specificity for the prediction of cycle cancellation in assisted reproduction. Ovarian volume, at a cut-off value 3mL$^3$, had high specificity for the prediction of non-pregnancy and cycle cancellation in assisted reproduction. Doppler studies of ovarian stromal blood flow were promising, but more research is needed.

**CRD commentary**

The review addressed a clearly stated research question, defined by appropriate inclusion criteria. The search strategy was wide ranging, included possible sources of unpublished studies and applied no language restrictions. Measures were taken to minimise error and/or bias when selecting studies for inclusion in the review and extracting data.

The authors stated that they extracted data on study quality, but no further details of this assessment, or results relating to study quality, were reported. The limited meta-analyses presented were broadly appropriate, given the heterogeneity of the data set. However, the lack of any results for individual studies severely limited interpretation.

Overall, the authors’ cautious conclusions accurately reflect the data presented. However, it might also have been said
that antral follicle count appeared to have high specificity for predicting non-pregnancy; a fuller presentation of the available data would have aided interpretation.

**Implications of the review for practice and research**

**Practice:** The authors stated that, because there were no studies using live birth rate as a final outcome (reference standard), it is difficult to make a strong case for routine use of antral follicle count or ovarian volume.

**Research:** The authors stated that further studies are needed to assess the performance of ovarian stromal blood flow, to determine the optimum parameters and the best cut-off values for all ultrasound based tests, and to determine if combinations of the three ultrasound tests (antral follicle count, ovarian volume and ovarian stromal blood flow), either in series or in parallel, could improve the predictability of fertility outcomes.

**Funding**

University of Aberdeen.

**Bibliographic details**


**PubMedID**

19802960

**DOI**

10.1080/14647270902896256

**Original Paper URL**

http://informahealthcare.com/doi/abs/10.1080/14647270902896256

**Indexing Status**

Subject indexing assigned by NLM

**MeSH**

Female; Fertility /physiology; Humans; Ovarian Follicle /blood supply /physiology /ultrasonography; Predictive Value of Tests

**AccessionNumber**

12009108023

**Date bibliographic record published**

02/12/2009

**Date abstract record published**

28/04/2010

**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.