Dynamic tests of ovarian reserve: a systematic review of diagnostic accuracy

Maheshwari A, Gibreel A, Bhattacharya S, Johnson NP

CRD summary
This review concluded that, based on the present level of evidence, dynamic tests for ovarian reserve should be abandoned completely. The review suffered from a number of limitations and the conclusions were not supported by the results of the review and are unlikely to be reliable.

Authors' objectives
To evaluate the accuracy of dynamic tests for ovarian reserve.

Searching
MEDLINE via PubMed, EMBASE and EBM Review (all from 1980) and CAB Abstracts (from 1990) and The Cochrane Library were searched to April 2008. Search terms were reported. Reference lists of primary studies and reviews were screened. No language restrictions were applied. Studies available only as abstracts were excluded.

Study selection
Studies that evaluated the accuracy of any dynamic test of ovarian reserve and in which the reference standard was fertility outcome were eligible for inclusion. Studies were excluded if they did not provide sufficient data to construct a 2x2 table of test performance, as were those that excluded women with abnormal test results. Included studies were both prospective and retrospective. Specific tests assessed were clomiphene citrate challenge test (CCCT), gonadotrophin-releasing hormone agonist stimulation test (GAST) and exogenous FSH ovarian reserve test (EFORT). Women in the included studies mostly underwent fertility treatment or had problems related to fertility; one study enrolled healthy women. Reference standards were pregnancy per cycle or per woman, poor response/cancelled cycle and recurrent pregnancy loss. Definition of abnormal test results varied across studies.

Two reviewers independently assessed studies for inclusion.

Assessment of study quality
Two reviewers independently assessed study quality based on the following criteria: study design; period of recruitment; relevant features of the population; test; reference standard; and outcomes of the study. Studies were considered to be of good quality if they enrolled consecutive women prospectively, had full verification of the test result with the reference standard and had adequate test description.

Data extraction
Two reviewers independently extracted data as 2x2 tables of test performance.

Methods of synthesis
Individual study estimates of sensitivity and specificity were plotted in receiver operating characteristic space. Summary sensitivity, specificity, positive and negative likelihood ratios (LR+, LR-) and diagnostic odds ratios (DOR) together with their 95% confidence intervals (CIs) were estimated for each test. The correlation between sensitivity and specificity was assessed using the Spearman correlation test.

Results of the review
Twenty studies (n=3,170) were included in the review.

Clomiphene citrate challenge test (CCCT) (20 studies, n=3,170): Summary sensitivity was 27% (95% CI 24% to 30%) and summary specificity was 92% (95% CI 89% to 94%) in the seven studies that used similar thresholds to define a positive test result (FSH>10 IU/l on day three or 10). Estimates were similar when the analysis was restricted to the five studies that enrolled women who were undergoing assisted reproduction treatment. Other thresholds and populations were each assessed in one or two studies.
GnRH agonist stimulation test (GAST) (12 studies, n=1,179): There was insufficient data to be combined across studies.

Exogenous FSH ovarian reserve test (EFORT) (seven studies, n=430): There was insufficient data to be combined across studies.

**Authors' conclusions**
The review was limited by heterogeneity between studies. Given the present level of evidence, these tests should be completely abandoned.

**CRD commentary**
This review addressed a clear objective. Inclusion criteria were defined only in terms of index test and reference standard. It appeared that studies had to report sufficient data to construct a 2x2 table of test performance to be included, but this was not explicitly defined as an inclusion criteria. The literature search was adequate for published studies, but no specific attempts were made to locate unpublished studies and so there was a possibility of publication bias. Study quality was assessed with appropriate criteria, but details on study quality were not reported or considered in the analysis. Appropriate steps were taken to minimise bias and errors in the selection of studies and extraction of data; it was unclear whether such steps were also taken for quality assessment. There were some limitations with the analysis. It is not appropriate to pool very heterogeneous studies, but the authors were very restrictive in data that were pooled. It may have been more helpful to have pooled data across a broader range of studies and then conducted subgroup analyses to investigate differences between studies. In particular, an analysis based on receiver operating characteristic (ROC) curves would have allowed pooling of data from studies that reported results for different thresholds. This would have allowed data from a larger number of studies to have contributed to the analysis. It would also have been helpful to have provided a narrative summary of results for tests that were considered inappropriate for meta-analysis. The authors' conclusions are unlikely to be reliable.

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

**Practice:** The authors stated that dynamic tests of ovarian reserve should be abandoned completely.

**Research:** The authors stated that there was an urgent need for consensus on the performance of dynamic tests for ovarian reserve and the definition of normality.

**Bibliographic details**

**PubMedID**
19549453

**Original Paper URL**
http://www.rbmojournal.com/article/S1472-6483(10)60019-3/abstract

**Indexing Status**
Subject indexing assigned by NLM

**MeSH**
Clomiphene /metabolism; Female; Follicle Stimulating Hormone /metabolism; Gonadotropin-Releasing Hormone /agonists; Humans; Oocytes /cytology /physiology; Ovarian Function Tests /methods; Ovary /cytology /physiology; Predictive Value of Tests; Pregnancy; Pregnancy Outcome

**AccessionNumber**
12009105806
Date bibliographic record published
05/08/2009

Date abstract record published
24/02/2010

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.