Diagnostic efficiency of home pregnancy test kits: a meta-analysis

Bastian L A, Nanda K, Hasselblad V, Simel D L

Authors' objectives
To assess the diagnostic efficacy of home pregnancy test (HPT) kits.

Searching
MEDLINE and HealthSTAR were searched for English language articles published between 1966 and 1996 using the keywords 'pregnancy', 'diagnosis', 'pregnancy test' and 'home test'. References cited in articles and those listed in the bibliographies of standard obstetric texts were also retrieved. Attempts were made to obtain information from the manufacturers of HPT kits, but none was obtained.

Study selection
Study designs of evaluations included in the review
The studies were required to include appropriate (non-pregnant) controls and to report on a sample size greater than 20. Review articles were excluded.

Specific interventions included in the review
No inclusion criteria were specified for the type of HPT included. The following types of pregnancy testing kits were included: Ova 11, Predictor, Answer, Daisy 2, e.p.t., e.p.t. plus, Advance, Fact, Answer 2, First Response and Acu-test. The kits were used both at home and in a laboratory. One study was excluded because the HPT used in it was no longer available.

Reference standard test against which the new test was compared
The included studies were required to compare HPT to a reference standard of laboratory testing. The laboratory test and cut-off point used were not specified in the inclusion criteria.

Participants included in the review
No inclusion criteria were specified with respect to the study participants. Testers of kits included the following two groups: volunteers who performed the pregnancy tests on study samples obtained previously by the investigators; and female patients who collected their own urine samples according to the kit instructions and performed the pregnancy tests on their own samples.

Outcomes assessed in the review
The included studies were required to report sufficient data to enable the calculation of sensitivity and specificity values. Where necessary, attempts were made to contact authors to obtain this information. The calculated sensitivity, specificity and test effectiveness scores were used as outcomes.

How were decisions on the relevance of primary studies made?
The articles retrieved were reviewed by two of the authors. The methods used to resolve any disagreements were not stated.

Assessment of study quality
The validity criteria were based on study design and level of evidence (see Other Publications of Related Interest). Two authors graded the studies A through C according to the validity criteria. The methods used to resolve any disagreements were not stated.

Data extraction
The data were extracted independently by two authors using a structured form. Any disagreements were resolved by
consensus. The sensitivity, specificity and test effectiveness scores were calculated.

Methods of synthesis
How were the studies combined?
Empirical Bayesian methods were used to estimate the summary sensitivity, specificity and test effectiveness scores with 95% confidence intervals (CIs). The summary estimates were calculated separately for studies where the participants were volunteers and those where the participants were female patients.

How were differences between studies investigated?
Tests for the homogeneity of sensitivity and test effectiveness were performed. No further details were given.

Results of the review
Five studies were included; 1 to 9 kits were tested per study, and the number of participants per kit tested ranged from 46 to 200. Of the 5 studies, 2 studies involved patients and 3 studies involved volunteers.

The included studies were graded A or B on methodological quality. Across kits and studies, the effectiveness of test score ranged from 0.09 to 3.7. The studies were found to be heterogeneous (P<0.05).

For volunteer participants, the summary sensitivity was 0.91 (95% CI: 0.84, 0.96) and the pooled test effectiveness was 2.75 (95% CI: 2.3, 3.2). For women who collected and tested their own urine, the summary sensitivity was 0.75 (95% CI: 0.64, 0.85) and the pooled test effectiveness was 0.82 (95% CI: 0.4, 1.2).

Forest plots of effectiveness scores with 95% CI were displayed for individual test kits and by patient or volunteer status.

Authors’ conclusions
The diagnostic efficiency of home pregnancy kits is greatly affected by the characteristics of the users. Despite the popularity of these kits, the relatively low effectiveness scores of these kits when used by actual patients are of concern. The authors suggest that the manufacturers of HPT kits publish results of trials in actual patients before marketing them to the general public.

CRD commentary
The aims of this review were clearly stated and the inclusion criteria were partially defined. Details of individual included studies were clearly reported. Although the use of a reference standard comparator is specified in the inclusion criteria, the nature of the reference standards used in the included studies is not defined and may represent a significant source of heterogeneity. There appears to have been inconsistency in the application of the inclusion criteria in that one study was excluded because ‘the kit was no longer available because of its demonstrated poor performance’; studies were included which related, either partially or wholly, to kits not currently on the market at the time of the review. The search was limited to published studies in the English language, therefore some relevant studies may have been omitted. A limited description was provided of the methods used to select the studies, grade validity and extract the data. Validity was assessed but details of the criteria used, though referenced, were not detailed in the review. It was reported that heterogeneity was assessed statistically, although details of the methods used were not given. While the data were stratified by participant type, no further investigation of heterogeneity was undertaken; the pooling of sensitivity values under these circumstances seems inappropriate. The method used to calculate ‘test effectiveness score’, though referenced, is not clearly described, and does not represent a standard method for describing overall test performance.

In view of the limitations highlighted, the authors’ conclusions should be treated with considerable caution.

Implications of the review for practice and research
Practice: Clinicians should be concerned about the diagnostic efficiency of HPT kits given the relatively low
effectiveness score when used by actual patients. When a patient calls reporting a negative result, she should be encouraged to repeat the test a week later if she remains amenorrheic and to call her health care provider if the test remains negative. HPT kit instructions should be reviewed to make sure women understand the instructions, to encourage women to wait at least 2 weeks after a missed period before performing the test, and to notify women of the potential for false-negative results.

Research: The authors suggested that further research on specific kits is required.

**Bibliographic details**


**PubMedID**

9755740

**Original Paper URL**


**Other publications of related interest**


**Indexing Status**

Subject indexing assigned by NLM

**MeSH**

Female; Humans; Predictive Value of Tests; Pregnancy; Pregnancy Tests /statistics & numerical data; Sensitivity and Specificity

**AccessionNumber**

11998001648

**Date bibliographic record published**

30/11/2003

**Date abstract record published**

30/11/2003

**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.