Authors' objectives
To evaluate the detection of Down's Syndrome using the triple-marker test.

Searching
MEDLINE was searched for articles published between January 1, 1966 and November 30, 1996 using the following keywords: 'Down syndrome-diagnosis', 'triple marker', 'triple-marker testing', 'triple screen', 'multiple-marker screening test' and 'prenatal diagnosis-methods'. Relevant gynaecological and medical journals were handsearched and all references in the retrieved studies were reviewed. No attempts were made to locate unpublished studies.

Study selection
Study designs of evaluations included in the review
Only diagnostic cohort studies written in English, French or German were eligible for inclusion.

Specific interventions included in the review
Studies that used the triple-marker test combining maternal serum alpha-fetoprotein (AFP), human chorionic gonadotrophin (hCG) and unconjugated estriol (uE3), and clearly defined a positive screening test, were eligible for inclusion.

Reference standard test against which the new test was compared
No inclusion criteria relating to the use of a specific reference standard test were defined. The included studies were required to define the reference standard used. The methods used to determine Down's syndrome included obtaining the foetal karyotype for all patients, and offering amniocentesis or chorionic villus sampling only to screen-positive patients. Information on the outcome of pregnancy (Down's syndrome or not) was obtained from sources including physicians' offices, cytogenetic laboratories and health centres. The determination of Down's syndrome status was not complete for all patients in all of the included studies.

Participants included in the review
No inclusion criteria relating to the study participants were specified. The participants in the included primary studies were pregnant women including those under and over the age of 35 years. Most had serum samples obtained between 15 and 22 weeks' gestation.

Outcomes assessed in the review
The included studies were required to report sufficient information to construct a 2x2 table. The outcomes calculated in the review were the sensitivity, false-positive rate and screen-positive rate of the test for Down's syndrome.

How were decisions on the relevance of primary studies made?
All of the eligible studies were reviewed independently by two unblinded authors, with any disagreements being resolved by discussion.

Assessment of study quality
The validity criteria included the following: selection of study participants; description of technique used; estimates of sensitivity, screen-positive rate and false-positive rate; cut-off used; blinding of the outcome assessments; follow-up of screened population; and accuracy estimated independently of the test threshold. Validity was assessed independently by two unblinded authors, with any disagreements being resolved by discussion.

Data extraction
The following data were extracted by two authors independently: author, year, sample size, maternal age (categorised as less than 35 years, 35 years and above, or all), the number of Down's syndrome cases, cut-off, sensitivity, false-positive rate and screen-positive rate.

**Methods of synthesis**

How were the studies combined?
The medians and ranges of sensitivities, false-positive rates and screen-positive rates were calculated. Summary receiver operator characteristic (sROC) curves were derived to compare and contrast the different point estimates of sensitivity and false-positive rates.

How were differences between studies investigated?
The Breslow-Day statistic was calculated. If significant heterogeneity was present (p<0.05) the studies were stratified according to maternal age (less than 35, 35 years and above, or all) and according to the cut-off values used (1:190 to 200; 1:250 to 295; and 1:350 to 380). Where there were sufficient data, the median and range of sensitivity, false-positive rate and screen-positive rate of the triple-marker screening test were calculated according to the cut-off values and the maternal age group.

**Results of the review**

Twenty-two cohort studies (194,326 patients) were included.

The Breslow-Day test (p<0.01) indicated significant heterogeneity across the studies. The cut-off values were 1:190 to 200 (6 studies), 1:250 to 295 (10 studies) and 1:350 to 380 (4 studies). Only one study used an ROC to estimate accuracy. All but one study adjusted the serum AFP values for maternal weight, race and the presence of insulin-dependent diabetes. The proportion of women with positive results that accepted prenatal diagnostic testing ranged from 67 to 92%. The median predictive values according to cut-off values were as follows.

Cut-off 1 (1:190 to 200): for 35 years and above, the sensitivity was 89% (range: 78 to 100), the false-positive rate was 25% (range: 20 to 29) and the screen-positive rate was 25% (range: 20.6 to 29.9). For all ages, the sensitivity was 67% (range: 48 to 91), the false-positive rate was 4% (range: 3 to 7) and the screen-positive rate was 4% (range: 3.6 to 6.0).

Cut-off 2 (1:250 to 295): for less than 35 years, the sensitivity was 57% (range: 53 to 58), the false-positive rate was 4% (range: 3 to 6) and the screen-positive rate was 4.4% (range: 3.2 to 6.0). For 35 years and above, the sensitivity was 80% (range: 75 to 100), the false-positive rate was 21% (range: 20 to 21) and the screen-positive rate was 20.9% (range: 20.6 to 21.0). For all ages, the sensitivity was 71% (range: 48 to 80), the false-positive rate was 6% (range: 4 to 7) and the screen-positive rate was 5.9% (range: 4.2 to 7.2).

Cut-off 3 (1:350 to 380): for all ages, the sensitivity was 73% (range: 70 to 80), the false-positive rate was 8% (range: 7 to 13) and the screen-positive rate was 8.2% (range: 8.2 to 13.5).

sROC curves in women of all ages: for a 5% false-positive rate, the sensitivity was 83.8% for cut-off 1 (1:190 to 200) and 57.3% for cut-off 2 (1:250 to 295). Insufficient numbers of studies assessed the test in women aged either above or below 35 years to obtain the ROC curve by age group.

**Authors’ conclusions**

Despite deficiencies in the study methods, triple-marker testing is an effective screening method for detecting Down syndrome pregnancies because it is simple, innocuous, relatively rapid, inexpensive, reproducible and noninvasive, with acceptable-to-high sensitivity and a low false-positive rate. The data indicate that the best cut-off value for predicting Down's syndrome is 1:190.

**CRD commentary**

This was a clearly written and presented review with a defined aim and inclusion criteria. Studies published in several languages were included. However, limiting the search to one database and published studies may have resulted in the
omission of relevant studies. More comprehensive details of the literature search, such as the titles of relevant gynaecological and medical journals searched, would have been welcome. The methods used to select the primary studies and assess validity were described, and the validity criteria were defined.

Heterogeneity was statistically assessed, some investigation was undertaken, and potential causes were considered. Further investigation of heterogeneity may have been helpful: for example, repeating the analysis after excluding studies without complete ascertainment of outcome and studies that did not adjust serum AFP for maternal weight, race and diabetic status; and stratifying the analysis by study validity. The use of sROC curves was appropriate, though the calculation of median sensitivity values for subgroups of possibly heterogeneous studies was of dubious value; division of a heterogeneous group of studies by arbitrary subgroup cannot be assumed to produce homogeneous subgroups.

The authors discussed several methodological problems with the primary studies: a lack of reporting the maternal age or proportion of women in the differing age groups; incomplete ascertainment of the outcome; the number of Down's syndrome cases based on the theoretical number calculated from the population screened; the lack of use of the best outcome reference standard; incomplete follow-up resulting in a potential underestimate of the number of Down's syndrome cases; variability in technical details of the test from one setting to another; different sample collection, storage and preparation; and differences in the protocols used for recalculating the risk of Down's syndrome when gestational age was corrected according to ultrasonic measures.

The authors' conclusions may be open to debate since high sensitivities were accompanied by high false-positive rates, with attendant implications for rates of unnecessary invasive testing and generation of patient anxiety. Low false-positive rates were only achieved where sensitivities were reduced to borderline levels of acceptability. Given the limitations outlined, it is difficult to draw firm conclusions on the value of the triple test as a screening tool from the data presented in this review.

Implications of the review for practice and research
Practice: The authors consider that women aged 35 years or older need to understand that a triple-marker test is less sensitive, and has a lower positive predictive value and a higher false-positive rate than does a karyotype obtained by chorionic villus sampling or amniocentesis.

Research: The authors consider that prospective studies are required to investigate the efficacy of screening for trisomy 21 in the first trimester using maternal serum markers AFP, free B-hCG, uE3 and pregnancy-associated plasma protein A, and to evaluate inhibin A in combination with triple-marker testing in the second trimester.

Bibliographic details

PubMedID
9618713

Indexing Status
Subject indexing assigned by NLM

MeSH
Adult; Down Syndrome /diagnosis /epidemiology /prevention & control; False Positive Reactions; Female; Humans; Mass Screening; Maternal Age; Pregnancy; Pregnancy, High-Risk; Prenatal Diagnosis; Sensitivity and Specificity

AccessionNumber
11998003778

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.