The accuracy of risk scores in predicting preterm birth: a systematic review


CRD summary
This review assessed the accuracy of published risk scores in predicting pre-term birth. The authors concluded that antenatal scoring systems are poor at predicting pre-term spontaneous birth in late pregnancy, and that there is a need for better quality information from well-designed studies. This was a well-conducted review and the authors’ conclusions are likely to be robust.

Authors’ objectives
To assess the accuracy of published risk scores in predicting pre-term birth.

Searching
Studies published in any language were sought. The sources searched (from inception to 2002, unless stated otherwise) were MEDLINE, EMBASE, Pascal, BIOSIS Previews, the Cochrane Library (Issue 2, 2002), MEDION, the National Research Register (Issue 2, 2002), SciSearch, and conference papers (1973 to 2002). Further information about the search strategy was reported in another paper (see Other Publications of Related Interest). The reference lists of reviews and selected studies were screened.

Study selection
Study designs of evaluations included in the review
Diagnostic accuracy studies were eligible for inclusion. Studies that did not clearly withhold information about test results from clinicians making decisions were excluded.

Specific interventions included in the review
Studies of clinical risk scoring tests were eligible for inclusion. The included studies used twelve different risk scoring systems. The most commonly used system was Creasy’s (based on a modification of the Papiernik-Berkhauer system); others included Copeland and a variety of statistically based systems. Details of the criteria used in all scoring systems and the test thresholds used in the individual studies were reported.

Reference standard test against which the new test was compared
Studies using spontaneous pre-term birth as the reference standard were eligible for inclusion. Most of the studies used spontaneous pre-term birth before 37 weeks, but not before 34 weeks, as the reference standard.

Participants included in the review
Studies of pregnant women where gestation was known at spontaneous birth were eligible for inclusion. Most of the studies assessed risk in women in their first trimester of pregnancy.

Outcomes assessed in the review
The review assessed accuracy, primarily using positive and negative likelihood ratios (LRs).

How were decisions on the relevance of primary studies made?
Two reviewers independently screened abstracts, while three reviewers independently selected studies from full publications of those identified. One reviewer assessed the eligibility of studies published in Italian, Spanish or German. Any disagreements were resolved by consensus, or through recourse to a third author.

Assessment of study quality
Studies were assessed for study design (cohort or case-control, retrospective or prospective), consecutive recruitment, adequate description of the test, and blinding of the test score results. Studies meeting all of these criteria were
considered to be of high quality. Two reviewers independently assessed validity.

**Data extraction**
Three reviewers independently extracted the data. For each study, data were extracted and used to construct 2x2 tables. The sensitivity, specificity, positive and negative LRs, and corresponding 95% confidence intervals (CIs), were calculated. Accuracy data were extracted for reference standards using spontaneous birth before 37 weeks and for earlier gestations.

**Methods of synthesis**
How were the studies combined?
The intention was to pool homogeneous studies using meta-analyses. When heterogeneity prohibited this, the range of values for positive and negative LRs were reported, along with their 95% CIs.

How were differences between studies investigated?
Statistical heterogeneity was assessed by examining forest plots, and tested using the chi-squared test. The effect on the results of repetition of scoring was explored. Subgroup analyses of studies using Creasy's scoring system were performed, stratified by race of participants or weeks gestation, using a random-effects meta-analysis.

**Results of the review**
Nineteen diagnostic accuracy studies (n=67,390) were included.

None of the studies met all of the quality criteria. Seventeen of the 19 studies were cohort studies, 13 studies used a prospective design, 4 studies enrolled consecutive women, 18 studies adequately described the test, and 2 studies used blinding. The quality of the included studies was not thought to explain the heterogeneity seen between studies.

The positive LRs ranged from 1.0 (95% CI: 0.6, 1.4) to 38.8 (95% CI: 23.5, 63.9) and the negative LRs from 0.1 (95% CI: 0.02, 0.6) to 1.2 (95% CI: 0.9, 1.6). Serial scoring did not improve accuracy (P=0.48).

Creasy's scoring system was used in 8 studies. Accuracy varied considerably.

For the 3 highest quality studies, the positive LRs ranged from 1.0 (95% CI: 0.6, 1.4) to 7.1 (95% CI: 5.0, 10.2) and the negative LRs from 0.5 (95% CI: 0.3, 0.6) to 1.0 (95% CI: 0.9, 1.2). Statistically significant heterogeneity was found for all subgroup analyses.

**Authors' conclusions**
Antenatal scoring systems were poor at predicting pre-term spontaneous birth in late pregnancy. There is a need for better quality information from well-designed studies.

**CRD commentary**
The review question was clear in terms of the study design, index test, reference standard and participants. Several relevant sources were searched with no language restrictions, thus minimising the potential for publication and language bias. Methods were used to minimise bias in the study selection, validity assessment and data extraction processes. The validity of the studies was assessed using appropriate criteria.

There was adequate information on the included studies. The methods used to combine the studies were appropriate, given the wide range of values and presence of statistically significant heterogeneity. Potential sources of heterogeneity were explored. This was a well-conducted review and the authors' conclusions are likely to be robust.

**Implications of the review for practice and research**
Practice: The authors stated that no recommendations for clinical practice could be made.
Research: The authors stated that further well-designed studies are required. Future studies should use clinically important outcomes as the reference standard and use robust methods, including blinding and the consecutive enrolment of women.

**Funding**
WellBeing, charity grant number K2/00.

**Bibliographic details**

**PubMedID**
15203570

**DOI**
10.1080/01443610410001685439

**Other publications of related interest**

**Indexing Status**
Subject indexing assigned by NLM

**MeSH**
Female; Gestational Age; Humans; Obstetric Labor, Premature /diagnosis /etiology; Predictive Value of Tests; Pregnancy; Prenatal Diagnosis; Risk Assessment; Risk Factors

**AccessionNumber**
12004009925

**Date bibliographic record published**
30/06/2006

**Date abstract record published**
30/06/2006

**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.