A systematic review of test performance in screening for oral cancer and precancer

Downer M C, Moles D R, Palmer S, Speight P M

CRD summary
This review assessed the accuracy of screening for oral cancer and pre-cancer, and concluded that screening is generally accurate. These conclusions do not reflect the results, which showed generally high specificity but poorer sensitivity. The review also suffered from a number of limitations in terms of the literature search, lack of a quality assessment, and methods of analysis.

Authors' objectives
To determine the accuracy of screening for oral cancer and pre-cancer in primary care.

Searching
MEDLINE, EMBASE, Cancerlit, AMED, CINAHL, British Nursing Index, HMIC, DARE and the Cochrane Library were searched from 1980 to January 2002 for studies reported in English; the search terms were reported and included a diagnostic filter. Only the most current primary report of each study was included.

Study selection
Study designs of evaluations included in the review
Prospective clinical studies were eligible for inclusion in the review.

Specific interventions included in the review
Studies that assessed screening for cancer or pre-cancer, and which involved an examination of the oral mucosa and defined criteria for a positive screen, were included. The personnel carrying out the screening included GDPs (not defined but likely to be General Dental Practitioners), hospital and general dentists, and health workers. Screening programmes were both pilot and definitive. Recruitment procedures were by invitation, opportunistic or through case finding. In some studies the personnel were not specifically trained. Screening was carried out in dental surgeries, health centres, hospitals and the participants’ homes. Details of the conditions targeted were reported. In some studies only individuals whose test results had been verified using the ‘gold’ standard were included in the review.

Reference standard test against which the new test was compared
Studies that used, as the reference standard, examination by an expert of participants that screened positive were eligible for inclusion. At least a proportion of individuals with a negative index test result had to receive the reference standard. Reference standards were not reported clearly but appeared to be clinical examination by dental health professionals such as oral medicine specialists, specialist oral surgeons, physicians, university-based dentists and project dentists. Reference standard examinations were carried out in dental surgeries, health centres, hospitals, the participants’ homes and referral centres.

Participants included in the review
No inclusion criteria relating to the participants were specified, but the review appeared to have been restricted to studies that screened healthy individuals. The categories of people screened in the included studies were company HQ staff, eligible adults, general practice patients, hospital visitors, high-risk adults and general adult populations. The studies were carried out in England, Japan, India and Sri Lanka. All studies included both men and women aged 20 years or older. In the included studies, the lesion prevalence in verified patients ranged from 1.4 to 50.7%.

Outcomes assessed in the review
The studies had to report sufficient data to construct a 2x2 table of test performance to be included. The primary outcomes reported in the review were the sensitivity and specificity.

How were decisions on the relevance of primary studies made?
One reviewer reviewed titles and abstracts, and those that were deemed to be of no relevance were excluded from further consideration. The full text of all remaining articles was obtained for detailed review by two independent reviewers. Any disagreements were resolved through consensus.

**Assessment of study quality**
The authors did not state that they assessed validity. The use of examiner blinding was reported for each study.

**Data extraction**
The data were extracted on a data extraction sheet for each study. The authors did not state how many reviewers performed the extraction. For each study, diagnostic test results (sensitivity, specificity, positive and negative predictive values, and positive and negative likelihood ratios) were presented.

**Methods of synthesis**
How were the studies combined?
A summary receiver operating characteristic (ROC) model was used to pool the studies using a random-effects model. The pooled sensitivity and the associated specificity were reported, along with 95% confidence intervals (CIs).

How were differences between studies investigated?
The studies were categorised into two groups based on the country in which the study was conducted (England or Japan versus countries in South East Asia). A random-effects meta-regression was then used to determine whether there was any difference in test performance between the two groups.

**Results of the review**
Seven reports describing eight studies were included (n=11,895).

The sensitivity ranged from 60 to 97% and the specificity from 75 to 99%. Specificity was much higher in the three studies conducted in England, Japan and India (94 to 99%) than in the two studies conducted in Sri Lanka (75 and 81%). The pooled sensitivity was 85% and the corresponding value of specificity, based on the summary ROC curve, was 97% (95% CI: 93, 98); the 95% CI for sensitivity based on the ROC curve was 73, 92. There were substantial differences in target populations, demographic characteristics, study designs, specified target lesions, categories, numbers of people undertaking screening and providing the reference standard clinical examinations, and the amount of training received by the screeners.

**Authors’ conclusions**
The studies showed a high level of discriminatory ability and consistency in test performance, irrespective of their clinical heterogeneity.

**CRD commentary**
This review suffered from a number of methodological limitations. Inclusion criteria were defined with respect to the outcomes; criteria for the index test and reference standard were broad, and none were defined for the participants. The literature search involved several databases but a diagnostic filter, which was likely to have resulted in relevant publications being missed, was employed. In addition, the review was limited to studies in English and no attempts to identify unpublished studies were made; the results might therefore be subject to language and publication bias. Although the authors defined their selection process as a form of quality assessment, quality was not formally assessed and the selection process did not include sufficient quality items to be considered a quality assessment. Methods used to avoid bias in the review process were appropriate, where reported, but details were not reported for all stages of the review.

There were insufficient details of the methods used to pool the studies to judge whether these were appropriate. A
summary ROC analysis was undertaken, which could be considered an appropriate method for pooling studies, but it was unclear how estimates of sensitivity were pooled and why the CI for this value was taken from the summary ROC curve rather than estimated from the model used to pool sensitivity. In addition, the authors stated that a meta-regression was carried out but did not provide further details on how this was conducted. Given the very small number of included studies it is questionable whether such an analysis was appropriate. The authors' conclusions should be interpreted with extreme caution given the limitations outlined above, especially as they do not reflect the relatively poor sensitivity of the screening examination.

**Implications of the review for practice and research**
The authors did not state any implications for practice or further research.

**Funding**
NHS R&D Health Technology Assessment (HTA) Programme, grant number 99/46/02.

**Bibliographic details**

**PubMedID**
14747057

**Indexing Status**
Subject indexing assigned by NLM

**MeSH**
Area Under Curve; Humans; Mass Screening /standards; Mouth Neoplasms /diagnosis; Precancerous Conditions /diagnosis; Primary Health Care /methods; Quality Assurance, Health Care; Sensitivity and Specificity

**AccessionNumber**
12004000408

**Date bibliographic record published**
31/08/2007

**Date abstract record published**
31/08/2007

**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.