Cochrane Library

Cochrane Database of Systematic Reviews

Clinical assessment to screen for the detection of oral cavity cancer and potentially malignant disorders in apparently healthy adults (Review)

Walsh T, Liu JLY, Brocklehurst P, Glenny AM, Lingen M, Kerr AR, Ogden G, Warnakulasuriya S, Scully C

Walsh T, Liu JLY, Brocklehurst P, Glenny AM, Lingen M, Kerr AR, Ogden G, Warnakulasuriya S, Scully C.
Clinical assessment to screen for the detection of oral cavity cancer and potentially malignant disorders in apparently healthy adults.
Cochrane Database of Systematic Reviews 2013, Issue 11. Art. No.: CD010173.
DOI: 10.1002/14651858.CD010173.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Review]

Clinical assessment to screen for the detection of oral cavity cancer and potentially malignant disorders in apparently healthy adults

Tanya Walsh[1], Joseph LY Liu[2], Paul Brocklehurst[1], Anne-Marie Glenny[3], Mark Lingen[4], Alexander R Kerr[5], Graham Ogden[6], Saman Warnakulasuriya[7], Crispian Scully[8]

[1]School of Dentistry, The University of Manchester, Manchester, UK. [2]University of Dundee, Dental Health Services Research Unit, Scottish Dental Clinical Effectiveness Programme, NHS Education for Scotland, Dundee, UK. [3]Cochrane Oral Health Group, School of Dentistry, The University of Manchester, Manchester, UK. [4]Pritzker School of Medicine, Division of Biological Sciences, Department of Pathology, University of Chicago, Chicago, Illinois, USA. [5]Department of Oral and Maxillofacial Pathology, Radiology and Medicine, New York University College of Dentistry, New York, USA. [6]Division of Oral and Maxillofacial Clinical Sciences, School of Dentistry, University of Dundee, Dundee, UK. [7]Clinical and Diagnostic Sciences, King's College London, London, UK. [8]University College London, London, UK

Contact address: Tanya Walsh, School of Dentistry, The University of Manchester, Coupland III Building, Oxford Road, Manchester, M13 9PL, UK. tanya.walsh@manchester.ac.uk.

Editorial group: Cochrane Oral Health Group.
Publication status and date: New, published in Issue 11, 2013.

Citation: Walsh T, Liu JLY, Brocklehurst P, Glenny AM, Lingen M, Kerr AR, Ogden G, Warnakulasuriya S, Scully C. Clinical assessment to screen for the detection of oral cavity cancer and potentially malignant disorders in apparently healthy adults. *Cochrane Database of Systematic Reviews* 2013, Issue 11. Art. No.: CD010173. DOI: 10.1002/14651858.CD010173.pub2.

ABSTRACT

Background

The early detection and excision of potentially malignant disorders (PMD) of the lip and oral cavity that require intervention may reduce malignant transformations (though will not totally eliminate malignancy occurring), or if malignancy is detected during surveillance, there is some evidence that appropriate treatment may improve survival rates.

Objectives

To estimate the diagnostic accuracy of conventional oral examination (COE), vital rinsing, light-based detection, biomarkers and mouth self examination (MSE), used singly or in combination, for the early detection of PMD or cancer of the lip and oral cavity in apparently healthy adults.

Search methods

We searched MEDLINE (OVID) (1946 to April 2013) and four other electronic databases (the Cochrane Diagnostic Test Accuracy Studies Register, the Cochrane Oral Health Group's Trials Register, EMBASE (OVID), and MEDION) from inception to April 2013. The electronic databases were searched on 30 April 2013. There were no restrictions on language in the searches of the electronic databases. We conducted citation searches, and screened reference lists of included studies for additional references.

Selection criteria

We selected studies that reported the diagnostic test accuracy of any of the aforementioned tests in detecting PMD or cancer of the lip or oral cavity. Diagnosis of PMD or cancer was made by specialist clinicians or pathologists, or alternatively through follow-up.

Data collection and analysis

Two review authors independently screened titles and abstracts for relevance. Eligibility, data extraction and quality assessment were carried out by at least two authors independently and in duplicate. Studies were assessed for methodological quality using QUADAS-2. We reported the sensitivity and specificity of the included studies.

Main results

Thirteen studies, recruiting 68,362 participants, were included. These studies evaluated the diagnostic accuracy of COE (10 studies), MSE (two studies). One randomised controlled of test accuracy trial directly evaluated COE and vital rinsing. There were no eligible diagnostic accuracy studies evaluating light-based detection or blood or salivary sample analysis (which tests for the presence of bio-markers of PMD and oral cancer). Given the clinical heterogeneity of the included studies in terms of the participants recruited, setting, prevalence of target condition, the application of the index test and reference standard and the flow and timing of the process, the data could not be pooled. For COE (10 studies, 25,568 participants), prevalence in the diagnostic test accuracy sample ranged from 1% to 51%. For the eight studies with prevalence of 10% or lower, the sensitivity estimates were highly variable, and ranged from 0.50 (95% confidence interval (CI) 0.07 to 0.93) to 0.99 (95% CI 0.97 to 1.00) with uniform specificity estimates around 0.98 (95% CI 0.97 to 1.00). Estimates of sensitivity and specificity were 0.95 (95% CI 0.92 to 0.97) and 0.81 (95% CI 0.79 to 0.83) for one study with prevalence of 22% and 0.97 (95% CI 0.96 to 0.98) and 0.75 (95% CI 0.73 to 0.77) for one study with prevalence of 51%. Three studies were judged to be at low risk of bias overall; two were judged to be at high risk of bias resulting from the flow and timing domain; and for five studies the overall risk of bias was judged as unclear resulting from insufficient information to form a judgement for at least one of the four quality assessment domains. Applicability was of low concern overall for two studies; high concern overall for three studies due to high risk population, and unclear overall applicability for five studies. Estimates of sensitivity for MSE (two studies, 34,819 participants) were 0.18 (95% CI 0.13 to 0.24) and 0.33 (95% CI 0.10 to 0.65); specificity for MSE was 1.00 (95% CI 1.00 to 1.00) and 0.54 (95% CI 0.37 to 0.69). One study (7975 participants) directly compared COE with COE plus vital rinsing in a randomised controlled trial. This study found a higher detection rate for oral cavity cancer in the conventional oral examination plus vital rinsing adjunct trial arm.

Authors' conclusions

The prevalence of the target condition both between and within index tests varied considerably. For COE estimates of sensitivity over the range of prevalence levels varied widely. Observed estimates of specificity were more homogeneous. Index tests at a prevalence reported in the population (between 1% and 5%) were better at correctly classifying the absence of PMD or oral cavity cancer in disease-free individuals that classifying the presence in diseased individuals. Incorrectly classifying disease-free individuals as having the disease would have clinical and financial implications following inappropriate referral; incorrectly classifying individuals with the disease as disease-free will mean PMD or oral cavity cancer will only be diagnosed later when the disease will be more severe. General dental practitioners and dental care professionals should remain vigilant for signs of PMD and oral cancer whilst performing routine oral examinations in practice.

PLAIN LANGUAGE SUMMARY

The detection of oral cavity cancers and potentially malignant disorders in apparently healthy adults

Cancer of the mouth is a serious condition and only half of those that develop the disease manage to survive after five years. It is commonly preceded by visible lesions, which if identified early, can be treated and could result in simpler surgery and much better outcomes. As a result, there is a need to understand how good different types of tests are at the early detection of oral cancer and the lesions that precede it. The most common method is an oral visual inspection by a clinician, but other tests include the use of a blue 'dye', illumination with a special light and self examination by the individual. The review found a lot of variety in the ability of the different tests to differentiate between healthy mouths and non-referable lesions and more serious lesions or oral cancer. Overall, visual examination by a front-line health worker proved to be the best method. Between 59% and 99% of mouth cancers were detected, although sometimes normal tissue was mistaken for oral cancer. The remaining techniques examined were not as good at detecting mouth cancer and identified less than a third of cases.

SUMMARY OF FINDINGS

Summary of findings 1.

What is the performance of conventional oral examination for the detection of oral cavity cancer and potentially malignant disorders in apparently healthy adults?

Population: Oral cavity cancer or potentially malignant disorder symptom-free individuals screened opportunistically, or through an organised screening programme

Index test: Conventional oral examination

Target condition: Oral cavity cancer or potentially malignant disorder

Reference standard: Examination and clinical evaluation by a physician with specialist knowledge or training. Long-term follow-up was accepted as a suitable reference standard for those participants screened negative

Studies: Cross-sectional (consecutive sample) (9) or validation sample in a randomised controlled trial of screening intervention (1)

	No. of participants (studies)	Effect (95% CI)
Population: Individuals attending for opportunistic screening (2), organised screening programme (4), validation as part of an organised screening programme or randomised controlled trial (3), screening as part of a routine surveillance appointment (1) **Index test:** Conventional oral examination **Prevalence**: Range from 1.4% to 50.9%	25,568 (10) No pooled analysis	Range: Sensitivity 0.50 (95% CI 0.07 to 0.93) specificity 0.98 (95% CI 0.92 to 1.00) Sensitivity 0.99 (95% CI 0.97 to 1.00) specificity 0.99 (95% CI 0.99 to 0.99)

CI = confidence interval

Summary of findings 2.

What is the performance of mouth self examination for the detection of oral cavity cancer and potentially malignant disorders in apparently healthy adults?

Population: Oral cavity cancer or potentially malignant disorder symptom-free individuals screened through an organised screening programme

Index test: Mouth self examination

Target condition: Oral cavity cancer or potentially malignant disorder

Reference standard: Examination and clinical evaluation by a physician with specialist knowledge or training or trained health worker

Studies: Cross-sectional studies (or consecutive series) (2)

	No. of participants (studies)	Effect (95% CI)
Population: Individuals attending for organised screening programme (2) **Index test:** Mouth self examination **Prevalence**: 0.6% and 22.6%	34,819 (2) No pooled analysis	Sensitivity 0.18 (95% CI 0.13 to 0.24) specificity 1.00 (95% CI 1.00 to 1.00) Sensitivity 0.33 (95% CI 0.10 to 0.65) specificity 0.54 (95% CI 0.37 to 0.69)

CI = confidence interval

Summary of findings 3.

What is the performance of vital rinsing (Toluidine blue) as an adjunct to conventional oral examination compared to conventional oral examination alone?

Population: Oral cavity cancer or potentially malignant disorder symptom-free individuals with tobacco habits

Index test: Conventional oral examination plus vital rinsing (Toluidine blue) compared to conventional oral examination alone

Target condition: Oral pre-malignant lesions and malignant lesions

Reference standard: Biopsy. Long-term follow-up through the National Cancer Registry

Studies: RCT (1)

	No. of participants (studies)	Effect (95% CI)
Population: Individuals attending an organised screening programme **Study**: Direct RCT **Index tests:** Conventional oral examination plus vital rinsing (Toluidine blue) compared with conventional oral examination alone **Prevalence**: 4.6% and 4.4%	7975 (1)	Detection rate of oral pre-malignant lesions and malignant lesions after referral was 4.6% in conventional oral examination plus vital rinsing arm; 4.4% in conventional oral examination alone. Rate ratio of 1.05 (95% CI 0.74 to 1.41). Incidence rate of oral cancer ($x10^{-5}$) of 28 compared to 35.4. Relative incidence rate of 0.79 (95% CI 0.24 to 1.23) * Initial screen positive rate higher in the vital rinsing arm (9.5% and 8.3%)

CI = confidence interval; RCT = randomised controlled trial

BACKGROUND

Target condition being diagnosed

The target conditions of interest are oral cavity cancer and potentially malignant disorders (PMD) of the lip and oral cavity. PMD is a term used to describe a range of lesions that present in the mouth and have the potential for malignant transformation. These include: erythroplakia, non-homogeneous leukoplakia, erosive lichen planus, oral submucous fibrosis and actinic keratosis (van der Waal 2009; Warnakulasuriya 2007).

The natural history of oral cancer is not fully understood (Napier 2008; Scully 2009). Carcinogenesis is a complex disease process; not all oral cancers will be preceded by PMD and not all PMD undergoes malignant transformation. Erythroplakia, non-homogeneous leukoplakia, erosive lichen planus, oral submucous fibrosis and actinic keratosis are the most important PMDs (Warnakulasuriya 2007) to proceed to carcinoma. Oral leukoplakia is the most common form of PMD and is defined as "white plaques of questionable risk having excluded (other) known diseases or disorders that carry no increased risk for cancer" (Warnakulasuriya 2007). Between < 1% and 18% of oral leukoplakias undergo malignant transformation. The presence of epithelial dysplasia can help predict malignant development in oral leukoplakia but the process is not linear; some mild dysplastic lesions undergo malignant transformation, whilst some severe lesions resolve (Jaber 2003; Reibel 2003). Carcinoma can also develop from lesions in which epithelial dysplasia was not previously diagnosed (Jaber 2003; Reibel 2003). As a result, most authorities regard leukoplakia as a dynamic rather than a static process (Napier 2008). In contrast, PMDs that are red or predominantly red in colour (e.g. erythroplakia and erythroleukoplakias) undergo malignant transformation more readily (Mashberg 1988; Mashberg 1995; Scully 2009).

Estimates of malignant transformation rates (MTR) vary enormously, from site to site within the mouth, from population to population and from study to study (Napier 2008). The MTR of hospital-based surveys are consistently higher than community-based studies because of sampling bias. Petti 2003 calculated a global MTR of oral leukoplakia of 1.36% per year (95% confidence interval (CI) 0.69% to 2.03%) based on the prevalence of oral leukoplakia, but this far exceeds the numbers of actual cases of malignancy. Virtually all studies emphasize the chronicity of oral PMD, with an increasing tendency to malignant change in the first five years. For example, the incidence of oral squamous cell carcinoma (OSCC) arising from leukoplakia in Californians was greatest in the second year of follow-up (11 out of 45; 24%) (Silverman 1984). The proportion of PMD that will develop OSCC is uncertain but low; best estimates suggest a rate of less than 2% per annum (Napier 2008).

The early detection and management of potentially malignant disorders of the lip and oral cavity that require intervention may reduce malignant transformations (though will not totally eliminate malignancy occurring), or if malignancy is detected during surveillance, there is some evidence that appropriate treatment may improve survival rates (Brocklehurst 2010; van der Waal 2009; Warnakulasuriya 2007). However, Lodi 2006 investigated the effectiveness of different management strategies for oral leukoplakia and found a lack of evidence for surgical interventions, including laser therapy and cryotherapy. Vitamin A, retinoids, beta carotene, carotenoids, bleomycin, mixed tea and ketorolac have also been tried, but none of the treatments tested showed a benefit when compared with the placebo. Lodi et al concluded that there was no evidence of effective treatment in preventing the malignant transformation of leukoplakia (Lodi 2006). There is also debate in the literature about the impact "field change" (Holmstrup 2006; Holmstrup 2009). Holmstrup argues that even if early lesions are surgically removed, the risk of malignant change can remain as a result of the lesion representing a small area of a wider field of damaged mucosa (Holmstrup 2006; Holmstrup 2009).

Technologies to treat and manage oral cancer have progressed substantially, as shown by systematic reviews of randomised controlled trials of interventions (e.g. Bessell 2011; Furness 2011; Glenny 2010). Once frank malignancy has been detected, the traditional management of oral cancer is through surgery and radiotherapy. More recently, systemic chemotherapy has been included as part of the treatment regimen before or during radiotherapy. Surgery for the treatment of oral cancer is followed by exacting reconstructive surgery to restore form and function. Debilitating side effects can occur as a result of both the surgery and radiotherapy and chemotherapy, adversely affecting an individual's quality of life. The five-year survival following diagnosis has remained at around 50% for the past 30 years in most countries (Parkin 2001; Warnakulasuriya 2009). Recent US data show a statistically significant improvement among patients treated for oral squamous cell carcinoma from 55% in 1984 to 1986 to 60% in the 1996 to 2003 time frame (Jemal 2008). This is in marked contrast to the improved survival rates in many other cancers, such as those of the breast and the colon (Cancer Research UK), but may be explained at least in some part by the fact that oral cancer is more often diagnosed at a late stage of the disease, when prognosis is poorer and the risks of significant morbidity and mortality are substantially higher (Rogers 2009; Rusthoven 2010).

Index test(s)

Reviews of primary studies of diagnostic test accuracy in this area have identified a number of index tests which could be used as adjuncts to the conventional visual and tactile oral examination (COE) to improve earlier detection of lip and oral cavity cancer and PMD (Fedele 2009; Leston 2010; Lingen 2008; Patton 2008; Rethman 2010). These include:

- vital rinsing or staining (Toluidine blue, Tolonium chloride)
- light-based detection (such as ViziLite and ViziLite Plus, Microlux/DL, VELscope, Orascoptic DK, Identafi 3000)
- mouth self examination
- blood and saliva analyses.

Vital rinsing and oral cytology are long available adjuncts to a conventional oral examination (Leston 2010; Lingen 2008). Other tests such as light-based detection systems have become commercially available only more recently. Mouth self examination is a simple technique with world wide application. Blood analysis and saliva analysis are more novel tests at an early stage of evaluation. It is worth noting that for an index test to obtain the US Food and Drug Administration (FDA) 'clearance' (the term reserved for non-invasive devices) a demonstration of efficacy is not required, only a demonstration of safety.

Of the index tests listed above, vital rinsing, light-based detection, mouth self examination and blood and saliva analyses could be used as screening adjuncts to the COE (Additional Table 1).

Where access to general dental practitioners or general medical practitioners is limited, either as a result of geographical location or barriers to uptake of healthcare provision, screening using the index tests listed above could, in principle, be undertaken by trained healthcare workers; all have the potential to be used as adjuncts to the COE by healthcare workers or clinicians undertaking screening of the general population. Adding any one of the proposed index tests to the COE, the tests could have a triage role in detecting lesions of uncertain significance with referral where appropriate. For instance, traumatic keratoses are common, and referring each patient with a white patch to a specialist to undergo a scalpel biopsy is excessive, and incurs increased financial cost and patient worry, and potentially delays the more urgent referrals being seen. A non-invasive index test or combination of tests adjunctive to the COE that provides a frontline clinician with a high degree of accuracy would not only reduce the number of patients with benign disease being referred, but could avoid the need for invasive biopsy in patients testing negative.

A companion Cochrane systematic review evaluates the diagnostic accuracy of index tests in individuals presenting with clinically evident lesions (Liu 2012).

Clinical pathway

The standard process of screening apparently healthy adults for PMD and cancer of the lip or oral cavity is by a systematic and thorough visual inspection of the oral mucosa and palpation of the neck under normal (incandescent) light for lymphadenopathy. In most instances this is carried out by a frontline clinician opportunistically as part of a routine recall examination by a dentist. This conventional visual and tactile oral examination (COE) used can be conducted with the minimum of effort and distress to the individual (Additional Table 1). Screening can be carried out opportunistically, for instance when an individual presents to their dentist for a check-up, or as part of an organised screening programme. The COE is usually followed by referral for further investigation if this is deemed necessary. The form that further investigation takes is variable nationally and internationally; it could be an examination/biopsy by a specialist in oral medicine or oral surgery at a secondary or tertiary clinic.

Rationale

Oral cancer is a significant global health problem with increasing incidence and mortality rates (Ferlay 2010; Warnakulasuriya 2009). Cancer of the lip or oral cavity is a relatively common cancer worldwide, with an estimated 263,000 new cases and 127,000 deaths in 2008, and an increasing incidence in recent years (Ferlay 2010). There is wide geographic variation in disease incidence and mortality, with almost double the incidence in low-income and middle-income countries as in high-income countries, and a threefold increase in mortality. Tobacco use, alcohol consumption, betel quid chewing and low socio-economic status have traditionally been thought to be the most important risk factors of oral cancer (Conway 2008; Faggiano 1997; La Vecchia 1997; Macfarlane 1995; Ogden 2005). Men have had a higher incidence of oral cancer than women (Ferlay 2010), but this disparity can be explained by men having a higher exposure to the above risk factors (Freedman 2007). The gender difference has narrowed in recent decades from a ratio of five males to one female diagnosed with oral cancers in the 1960s to less than two to one in 2008 (Ferlay 2010). Although traditionally the risk of oral cancer increases with age, the incidence among younger adults has been increasing in the European Union and the United States (Warnakulasuriya 2009). In the United Kingdom, one in 10 cases are now below the age of 45 years (Cancer Research UK).The five-year survival rate depends on multiple factors, including patient and tumour characteristics, treatment received and stage at diagnosis. Oral cancer incidence and mortality can be reduced using three approaches: (i) primary prevention, (ii) secondary prevention, screening and early detection, and (iii) improved treatment (Scully 2000b).

Successful early detection of oral cavity cancer or PMD is highly dependent on whether individuals with the disease present for a screen. Early detection relies on the awareness and motivation of the clinician or patient in identifying a suspicious lesion or symptom while it is still at an early stage. Whilst many organisations advocate cancer-related checks, including the American Cancer Society for individuals of all risk groups (American Cancer Society 1992) and the US Preventive Health Services Task Force for high risk individuals (US Preventive Services Task Force in discussion), there is much global variation in the provision and promotion of routine oral cancer examinations. Currently, no national population-based screening programmes for oral cancer have been implemented in the high-income countries, although opportunistic screening has been advocated (Brocklehurst 2013). Consequently, individuals will often present for examination at a later stage of the disease, when the risks of significant morbidity and mortality are substantially higher. The British Columbia Oral Cancer Prevention Program (BC OCPP) is addressing this challenge in several ways: by linking community dental practices and referral centres, by creating partnerships between scientists and clinicians that already have resulted in new technologies to enhance early diagnosis, by involving a broad range of stakeholders to ensure population-based screening and by engaging in provincial, national and international outreach (Rosin 2006). Brocklehurst et al's systematic review identified only one randomised controlled trial using visual examination with a follow-up period of 15 years which was carried out in India. The authors of the review concluded that opportunistic screening of high risk groups may potentially improve outcomes, although the risk of bias of the included study was high (Brocklehurst 2013).

There is some debate in the literature on anticipated differences in diagnostic accuracy of prospective population-based invitational screening programmes and a more opportunistic approach (when patients attend their general (dental) practitioner for routine examination or for treatment). In Downer et al's systematic review of test performance in screening for oral cancer and PMD, only prospective investigations of population screening with specified reference standards were included. The pooled sensitivities and specificities were 0.85 (95% CI 0.730 to 0.919) and 0.97 (95% CI 0.930 to 0.982) respectively (Downer 2004). An opportunistic approach that focuses on high risk groups is also possible (McGurk 2010; Sankaranarayanan 1997). A simulation study which used neural network and machine learning techniques suggested opportunistic screening aimed at high risk groups may be both effective and cost-effective (Speight 2006). However, many individuals with risk factors may not attend the dentist and are therefore not amenable to an opportunistic approach (Netuveli 2006; Yusof 2006).

Reviews assessing the test accuracy of a conventional oral examination as a population screening tool (e.g. Downer 2004;

Moles 2002) have highlighted methodological flaws in the primary diagnostic test accuracy studies, although explicit methodological quality assessment of these studies using a validated and widely used checklist was not undertaken.

In this review we have identified screening tests for PMD and cancer of the lip or oral cavity to evaluate the diagnostic accuracy of the COE and the accuracy of the other index tests (Additional Table 1) used as adjuncts to the oral examination in asymptomatic adults. The index tests proposed for evaluation in this review are suitable for use in the community or as part of a dental examination in a general dental practitioners' office. The review includes both prospective investigations of organised screening programmes and prospective opportunistic screening. It is important that this review considered both as individuals screened opportunistically are self selecting and may not be representative of the population of interest. In either scenario, screening may be carried out by dental professionals or healthcare workers. The purpose of the screening is to identify the presence or absence of PMD which require referral to secondary care for definitive diagnosis and possibly treatment. The proposed index tests cannot confirm whether a PMD is cancerous before deciding on referral to secondary care; biopsy with histopathology is currently the only confirmatory method of oral cancer diagnosis.

The Cochrane Oral Health Group has undertaken a number of intervention reviews in the field of treatment of oral and oropharyngeal cancers (Bessell 2011; Furness 2011; Glenny 2010) and screening programmes for the early detection and prevention of oral cancer (Brocklehurst 2013). This screening test accuracy review complements the intervention reviews.

OBJECTIVES

To estimate the diagnostic accuracy of conventional oral examination (COE), vital rinsing, light-based detection, biomarkers and mouth self examination (MSE), used singly or in combination, for the early detection of potentially malignant disorders (PMD) or cancer of the lip and oral cavity in apparently healthy adults.

Secondary objectives

To estimate the accuracy of the different index tests with COE, when compared with each other.

METHODS

Criteria for considering studies for this review

Types of studies

We included studies of cohorts of apparently healthy adults which evaluated the diagnostic accuracy of the conventional oral examination (COE) used singly or in combination with an index test listed in Additional Table 1, in screening for potentially malignant disorders (PMD) and cancer of the lip or oral cavity. These included cross-sectional studies (or consecutive series) and randomised controlled trials (RCTs) of test accuracy. We excluded case series and case-control studies which could lead to inflated estimates of prevalence and test accuracy (Whiting 2004). We also excluded studies reported in abstract form alone, uncontrolled reports and randomised controlled trials of the effectiveness of screening programmes (intervention studies). Where randomised or paired comparative designs were available these were included in the review and analysed separately. Only studies reporting data for true positives, false positives, true negatives and false negatives at an individual level (as opposed to a lesion level) for each test were included. No language restrictions were imposed.

Participants

Apparently healthy adults not reporting symptoms attending an organised screening programme or screened during attendance at a dental or other clinical practice examination. We did not exclude specific subgroups of patients in this review, such as high risk cohorts or cohorts with previous suspicions on PMD or cancer of the lip or oral cavity.

Index tests

The COE used as a screen, alone or in combination with any other screening tests previously listed (Additional Table 1). The COE (conventional testing test) was the initial point of the screen, which all individuals received. The index test was used as an adjunct following the COE irrespective of whether oral cancer or PMD was suspected by the COE alone (i.e. a positive test result is a positive result from either the COE or the index test or both).

Target conditions

Following the consensus views of the expert working group of the World Health Organization (WHO) Collaborating Center for Oral Cancer and Precancer, the target conditions of the lip or oral cavity of interest are noted as.

- Carcinoma of the lip or oral cavity.

- Potentially malignant disorders.

- Leukoplakia.
- Erythroplakia.
- Lichen planus.
- Lupus erythematosus.
- Submucous fibrosis.
- Actinic keratosis.
- Hereditary disorders such as dyskeratosis congenita or epidermolysis bullosa.

Reference standards

The reference standard was examination and clinical evaluation by a physician with specialist knowledge or training, working to the current diagnostic guidelines of their locality. At the most experienced level this would be an oral and maxillofacial pathologist or oral medicine specialist possibly utilising biopsy with histology where considered clinically appropriate. More commonly this was expected to include general dental physicians in receipt of supplementary training in the detection and identification of PMD and carcinoma of the lip or oral cavity or other physicians with dedicated training. We included studies where confirmation of individuals screened negative by the index test was done by extended follow-up. To be eligible for inclusion in the review, at least a proportion of the screened negatives were required to be verified. Where reported, for each study we noted the diagnostic protocol, guidelines or registry used for follow-up in the Characteristics of included studies table. Studies with confirmatory biopsy of individuals who screened negative by the index test

were eligible for inclusion although ethically questionable (Downer 2004).

Search methods for identification of studies

Electronic searches

We searched the following electronic databases.

- The Cochrane Diagnostic Test Accuracy Studies Register (to 30 April 2013).
- The Cochrane Oral Health Group's Trials Register (to 30 April 2013).
- MEDLINE via OVID (1946 to 30 April 2013).
- EMBASE via OVID (1980 to 30 April 2013).
- MEDION (2003 to 30 April 2013).

See Appendix 1 for the search strategies used. There were no restrictrions on language in the searches of the electronic databases.

We constructed the electronic search strategy in accordance with this review and that of a companion Cochrane diagnostic test accuracy review (Liu 2012) which was undertaken concurrently by the same review team.

Searching other resources

We sought to locate further studies through citation searches and reference lists of key articles, and by contacting authors of identified articles to request information of any unpublished or ongoing studies.

Data collection and analysis

Selection of studies

We did not limit the screening of the search results by publication language or status. Non-English articles were translated. Titles and abstracts of all articles identified from the searches were independently assessed by two review authors. For articles appearing to meet the inclusion criteria, or where a clear decision was unable to be made from scanning the title and abstract alone, full reports were obtained. Where disagreements occurred, these were resolved by discussion with the review team.

Data extraction and management

Two review authors independently extracted data using a piloted data collection form. Discrepancies were resolved through discussion with the review team. Study authors were contacted to obtain relevant missing data if these were not available in the printed report.

From each study, we extracted the following data.

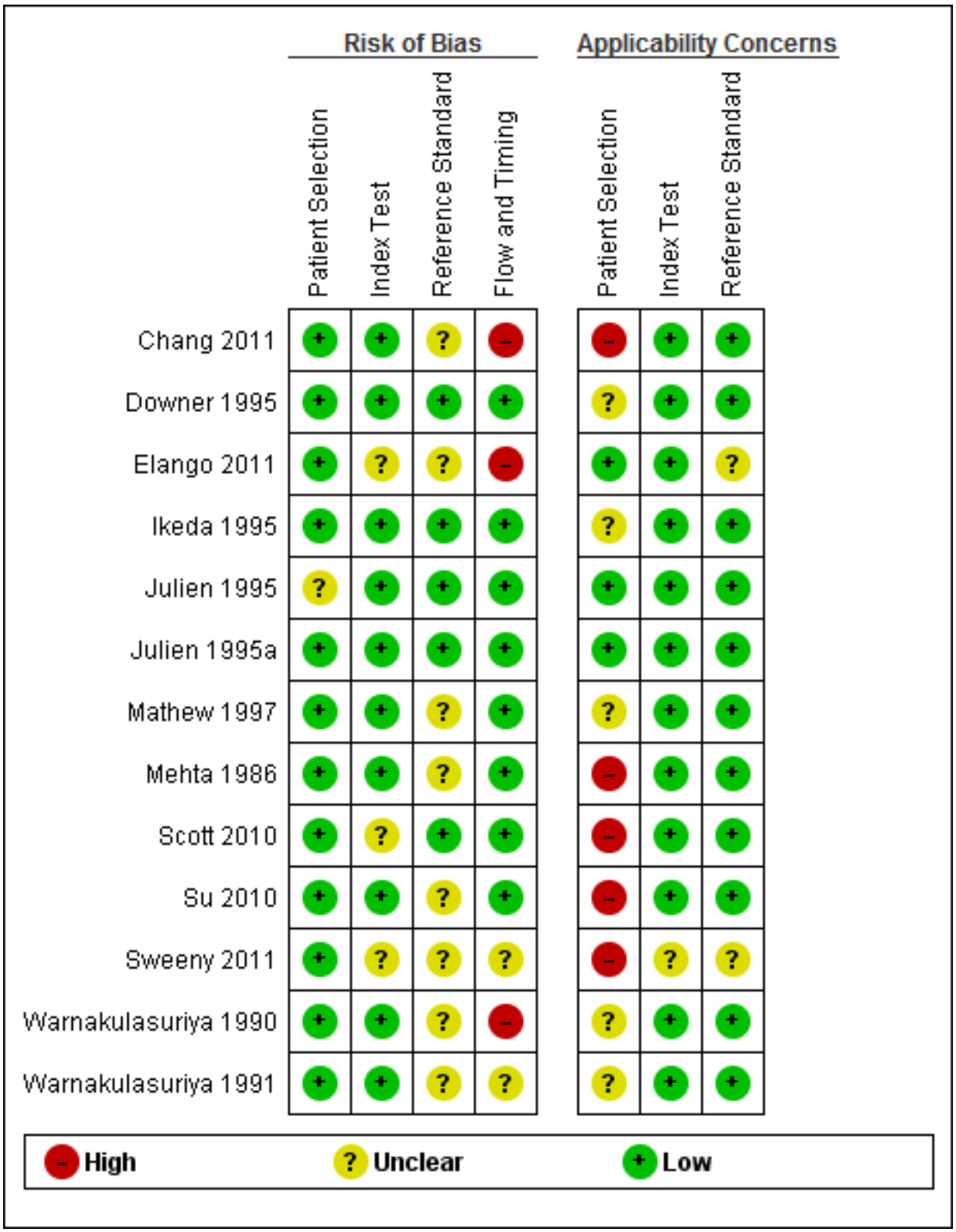
- Sample characteristics (age, sex, socio-economic status, risk factors (e.g. human papillomavirus (HPV) status, prevalence of tobacco use and alcohol consumption), number of participants/ lesions).
- Setting (country, disease prevalence, type of screening).
- Type of index test(s) (category, name, positivity threshold).
- Study information (design, reference standard, case definition, training and calibration of personnel).
- Study results (true positive, true negative, false positive, false negative, any equivocal results, withdrawal or exclusions).

This information was documented in the Characteristics of included studies table for each study.

Assessment of methodological quality

We used the QUADAS-2 tool (Whiting 2011) to assess the quality of the included studies over four key domains: patient selection, index test, reference standard and flow and timing of participants through the study. The QUADAS-2 tool was tailored specifically for this review (Additional Table 2). Review specific guidance was used to facilitate documentation of the pertinent descriptive information contained in the studies. Customised instructions to aid judgement of the signalling questions were given (following Patton 2008). Two core signalling questions were removed: 'Was a case-control design avoided?' (this study design was excluded from the review); 'Did all patients receive a reference standard?' (this was a criterion for inclusion). Two additional signalling items relating to commercial funding and multiple index tests were added to the core signalling questions. Responses to the signalling questions, risk of bias and applicability judgements are presented in the Characteristics of included studies tables and summarised graphically (Figure 1).

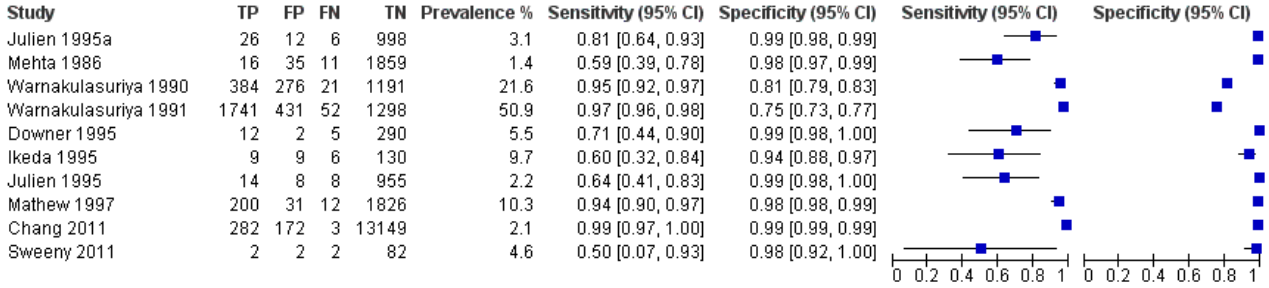
Figure 1. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study.



Statistical analysis and data synthesis

Data for the true positive, true negative, false positive and false negative values for each test in each study was entered into Review Manager (RevMan 2012). For each index test, estimates of the diagnostic accuracy were expressed as sensitivity and specificity with 95% confidence intervals. This information was displayed as coupled forest plots (Figure 2), and plotted in receiver operating characteristic (ROC) space.

Figure 2. Forest plot of 1. Conventional oral examination.



For the primary analysis we had intended to undertake a meta-analysis to combine the results of the studies for each index test. However, the substantial diversity of characteristics of the included studies meant that this was not appropriate.

We were only able to include one study (Su 2010) that directly evaluated the comparative accuracy of more than one index test with the reference standard, i.e. randomising individuals to different index tests. This study was reported separately.

Investigations of heterogeneity

We planned to explore possible sources of heterogeneity through meta-regression including the following covariates: characteristics of the study sample (prevalence of carcinoma or PMD in the study (> 50% prevalence); inclusion of HPV + adults; tobacco users/ high alcohol consumption); target condition (oral squamous cell carcinoma alone or oral squamous cell carcinoma and PMD); aspects of study design (prospective organised or opportunistic); type of reference standard (examination and clinical evaluation by physician with specialist knowledge or extended follow-up) and operator (dental or general medical practice professionals or other healthcare workers). Given the diversity of the studies this was not undertaken.

Sensitivity analyses

No sensitivity analyses were undertaken.

Assessment of reporting bias

Tests for reporting bias were not conducted because current tests are misleading when applied to systematic reviews of diagnostic test accuracy (Leeflang 2008).

RESULTS

Results of the search

After de-duplication the initial electronic search conducted in April 2013 retrieved 4220 records.These were screened independently and in duplicate according to eligibility criteria; 33 records were considered potentially eligible for inclusion. Of this number, 17 records with 13 studies were included in the review. The main reasons for exclusion were ineligible study design or no reference standard data for individuals screened negative. Ten studies reported on the diagnostic accuracy of conventional oral examination (COE) alone; two studies reported on mouth self examination and one randomised controlled trial directly compared COE alone with COE plus a vital rinsing agent (Toluidine blue). No diagnostic test accuracy studies meeting the review inclusion criteria evaluating any other pre-specified index test were found.

Four studies are still awaiting classification and one is ongoing.

Methodological quality of included studies

The assessment of methodological quality is presented graphically in Figure 1.

Conventional oral examination

The nature of the screening of participants can be broadly categorized into opportunistic screening (Chang 2011; Julien 1995), organised screening programmes (Downer 1995; Jullien 1995a; Warnakulasuriya 1990; Warnakulasuriya 1991), validation as part of an organised screening programme or randomised controlled trial (Ikeda 1995; Mathew 1997; Mehta 1986) and screening as part of a routine surveillance appointment (Sweeny 2011).

The accuracy of detecting potentially malignant disorders (PMDs) and oral cavity cancer was evaluated in a variety of different settings: In Tokoname, Japan, all residents of 60 years of age were invited by mail to attend a dental screening programme at a health centre (Ikeda 1995). In Kerala, India, basic healthcare workers incorporated screening into their routine house visits (Mathew 1997; Mehta 1986) as in Sri Lanka (Warnakulasuriya 1990; Warnakulasuriya 1991). In the United Kingdom, the feasibility and accuracy of workplace screening was evaluated in one study (Downer 1995), of screening patients at a medical practice in another (Julien 1995a), and opportunistically in patients attending a dental hospital for an out-patient appointment (Julien 1995). In Taiwan, screening was offered to individuals attending a tertiary referral centre (Chang 2011). In the USA, screening was part of the routine surveillance visit of patients attending an otolaryngology clinic (Sweeny 2011).

Risk of bias for the patient selection domain was low for all studies with one exception (Julien 1995). This study was judged as unclear as the method of patient selection for this opportunistic screening study was not reported. Two studies were judged to be of low concern for applicability (Julien 1995; Julien 1995a); five studies of unclear applicability as a result of not fully reporting the participant characteristics or risk factors of the study sample or both (Downer 1995; Ikeda 1995; Mathew 1997; Warnakulasuriya 1990; Warnakulasuriya 1991). Three studies were selective in their sampling, targeting a 'high risk' population. These were all male patients attending the otolaryngology or dental department (Chang 2011), previous cancer patients attending the otolaryngology clinic for a routine surveillance visit (Sweeny 2011) and individuals over 35 years of age with "tobacco habits" (Mehta 1986).

The COE index test was carried out by clinicians (general dental practitioners, community dental officers, otolaryngologists) in six studies (Chang 2011; Downer 1995; Ikeda 1995; Julien 1995; Julien 1995a; Sweeny 2011) and by health workers in the studies in India and Sri Lanka (Mathew 1997; Mehta 1986; Warnakulasuriya 1990; Warnakulasuriya 1991). The risk of bias judgements for this domain were judged to be low in nine studies. The index test was carried out prior to the reference standard and a positivity threshold for the target condition was specified a priori. One study (Sweeny 2011) was judged to be at unclear risk of bias as there was a lack of clear definition of the target condition and the positivity threshold. All studies were judged to be at low concern regarding applicability.

Four studies (Downer 1995; Ikeda 1995; Julien 1995; Julien 1995a) were judged to be at low risk of bias for the reference standard domain. In these studies the reference standard was carried out by experienced specialist physicians and the results were interpreted without knowledge of the results of the index tests. For the remaining studies it was unclear whether the reference standard personnel were unaware of the results of the index test when interpreting the reference standard. One study (Sweeny 2011) was judged to be at unclear concern regarding applicability as the target definition was recurrence of head and neck cancer; all other studies were judged as low concern.

For the flow and timing domain, two studies were judged to be at high risk of bias as a result of attrition following positive screen (37.5% of screen positive) and differential verification (Chang 2011) and time from screen positive to receiving reference standard (Warnakulasuriya 1990). Two studies were judged to be at unclear risk of bias (Sweeny 2011; Warnakulasuriya 1991), the remainder at low risk of bias (Downer 1995; Ikeda 1995; Julien 1995; Julien 1995a; Mathew 1997; Mehta 1986).

Two studies (Chang 2011; Warnakulasuriya 1990) were judged as being at overall high risk of bias resulting from the flow and timing domain; three studies were at overall low risk of bias (Downer 1995; Ikeda 1995; Julien 1995a). For the remaining five studies an unclear risk of bias for at least one of the four domains resulted in an overall risk of bias judgement of unclear (Julien 1995; Mathew 1997; Mehta 1986; Sweeny 2011; Warnakulasuriya 1991).

Three studies (Chang 2011; Mehta 1986; Sweeny 2011) were judged as having high overall concerns regarding applicability, arising from patient selection of high-risk groups. Two studies (Julien 1995; Julien 1995a) were judged as having low overall concerns regarding applicability. For the remaining five studies an unclear concern regarding applicability in the patient selection domain resulted in an overall applicability judgement of unclear (Downer 1995; Ikeda 1995; Mathew 1997; Warnakulasuriya 1990; Warnakulasuriya 1991).

Mouth self examination

Two studies (Elango 2011; Scott 2010) evaluated mouth self examination as part of an organised screening programme. Risk of bias for patient selection was judged to be low for both studies. Concerns regarding applicability for this domain were judged as low for one study (Elango 2011) and high for the other (Scott 2010). In this study, the study sample consisted of participants older than 45 years of age with tobacco habits.

We gave a judgement of unclear risk of bias to both studies for the index test domain, as it was not reported whether the results of the index test were interpreted without knowledge of the reference test. We gave a judgement of low concerns regarding applicability for this domain.

The risk of bias judgement for the reference standard domain was low for one study (Scott 2010), being evaluated by a dentist with training and the reference test being carried out prior to the index test. We judged the other study (Elango 2011) as unclear risk of bias as it was unclear whether the conduct of the reference standard would be likely to correctly classify the condition and also whether the reference standards were interpreted without knowledge of the index test. The manuscript states that "the competence of the health workers [reference standard] was confirmed by a trained oral cancer specialist" but not reported. Consequently the judgements of concerns regarding applicability for this domain were low (Scott 2010) and unclear (Elango 2011).

Risk of bias was judged to be low for the flow and timing domain (Scott 2010) and high (Elango 2011) due to a significant number of withdrawals and exclusions for non-compliance.

The overall risk of bias was judged to be unclear (Scott 2010) and high (Elango 2011). Concern regarding the overall applicability of the studies to the review question was high (Scott 2010) due to patient selection and unclear (Elango 2011) due to the reference standard being carried out by general health workers specifically trained for the study rather than a specialist or experienced clinician.

Conventional oral examination compared to conventional oral examination plus vital rinsing (Toluidine blue)

We judged this study (Su 2010) which directly compared two index tests in a randomised controlled trial to be at low risk of bias for patient selection and index test. Concerns regarding applicability were judged as high for the patient selection domain as individuals who "lacked oral habits" such as smoking or betel quid chewing were eligible for the trial. We judged that there were low concerns regarding applicability of the index tests. We judged the study to be at unclear risk of bias whether this was interpreted without knowledge of the results of the index tests is unclear. There was low concern regarding applicability of the reference standard. Risk of bias for the flow and timing domain was judged as low.

Overall risk of bias for this study was judged as unclear, based on the interpretation of the reference standard. Concern regarding the overall applicability of the study was high, arising from patient selection.

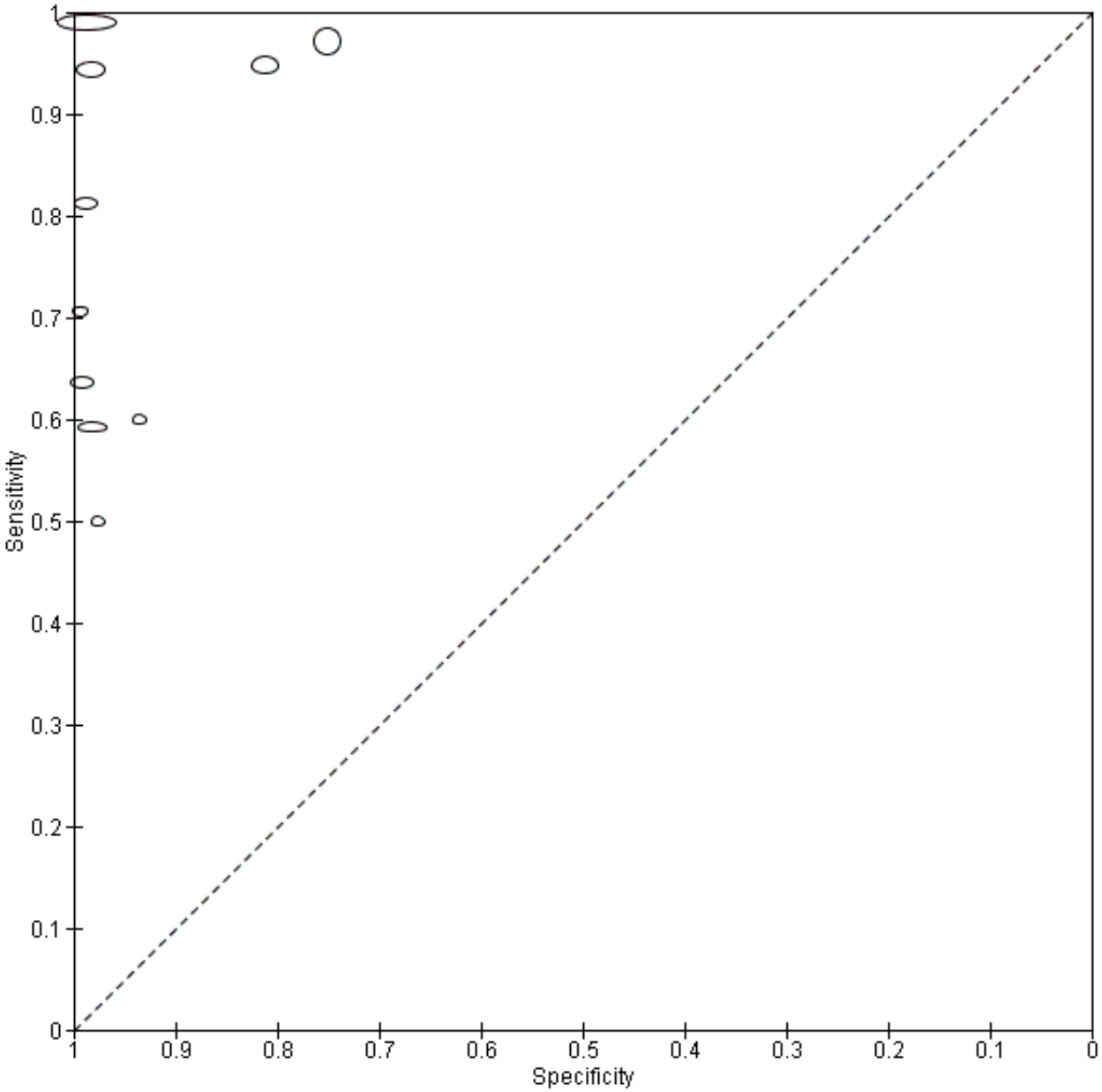
Findings

Conventional oral examination

Diagnostic accuracy of COE by a non-specialist compared to a reference standard was evaluated in 10 studies including 25,568 participants in total, where the target condition was PMD and cancer of the lip or oral cavity. Pooling of the studies was considered inappropriate due to the diversity of study and participant characteristics. The prevalence of PMD or oral cavity cancer in the diagnostic test accuracy study samples ranged from 1.4% to 50.9%. For the eight studies with prevalence of 10% or lower, the sensitivity estimates were highly variable, and ranged from 0.50 (95% confidence interval (CI) 0.07 to 0.93) to 0.99 (95% CI 0.97 to 1.00) with uniform specificity estimates around 0.98 (95% CI 0.97 to 1.00). Estimates of sensitivity and specificity were 0.95 (95% CI 0.92 to 0.97) and 0.81 (95% CI 0.79 to 0.83) for one study with prevalence of 21.6% and 0.97 (95% CI 0.96 to 0.98) and 0.75 (95% CI 0.73 to 0.77) for one study with prevalence of 51%.

Study prevalence is shown in the coupled forest plot (Figure 2) along with estimates of sensitivity and specificity and also plotted in ROC space (Figure 3). All studies for this index test used a common threshold, the presence of PMDs and oral cancer.

Figure 3. Summary ROC plot of 1. Conventional oral examination.



A summary is given in the Summary of findings 1.

Mouth self examination

Two studies (Elango 2011; Scott 2010) provided data from 34,819 individuals. The prevalence was very different in the two studies: 0.6% and 22.6% respectively. Values of sensitivity were low in both studies (0.18 (95% CI 0.13 to 0.24) (Elango 2011) and 0.33 (95% CI 0.10 to 0.65) (Scott 2010)) but values of specificity were higher (1.00 (95% CI 1.00 to 1.00) (Elango 2011) and 0.54 (95% CI 0.37 to 0.69) (Scott 2010)) (Figure 4; Figure 5).

Figure 4. Forest plot of 2. Mouth self examination.

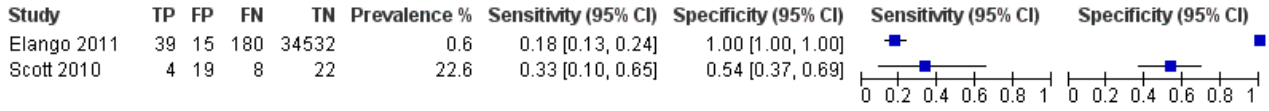
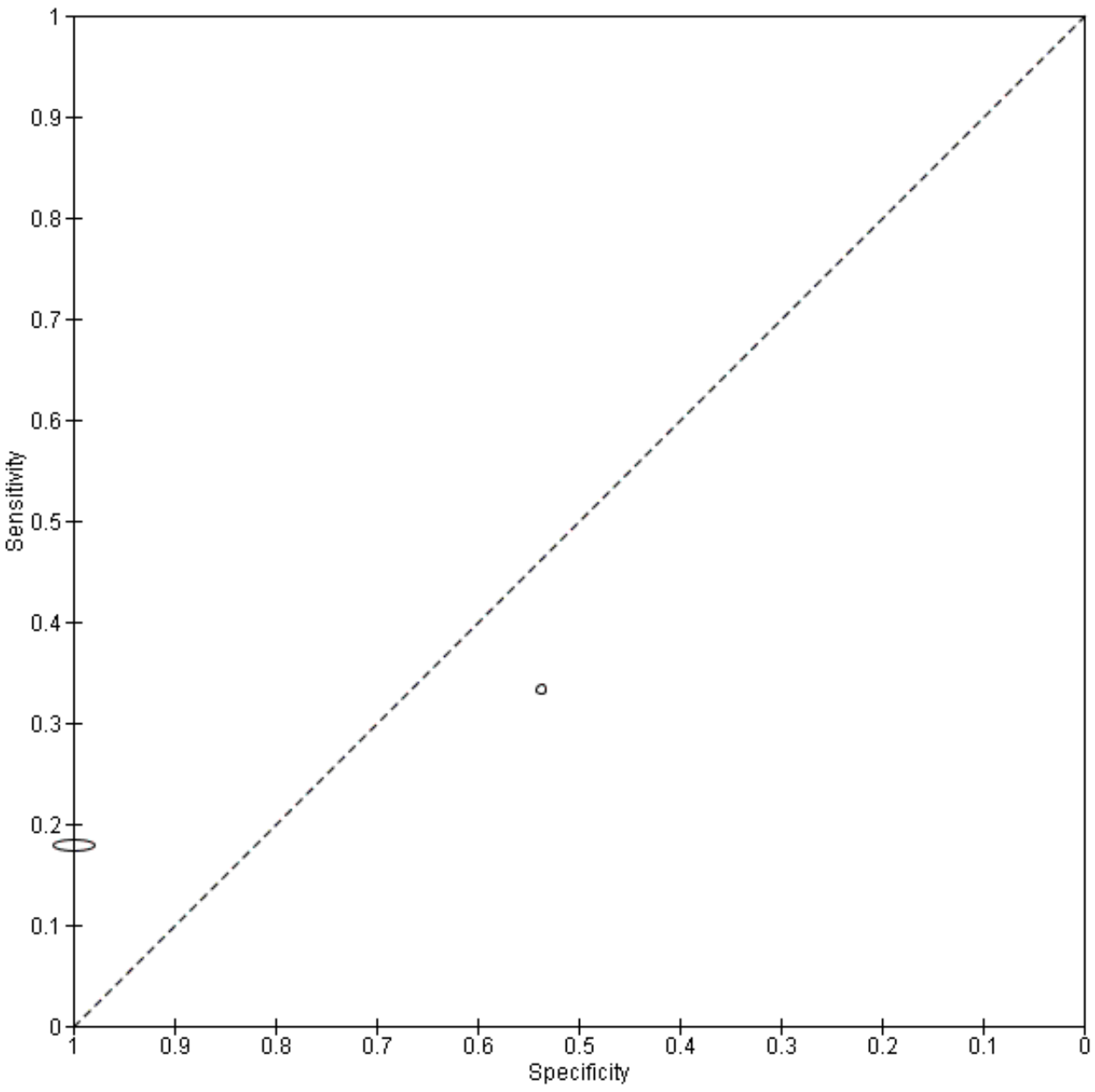


Figure 5. Summary ROC plot of 2. Mouth self examination.



A summary is given in the Summary of findings 2.

Conventional oral examination compared to conventional oral examination plus vital rinsing (Toluidine blue)

We included one randomised controlled trial which directly compared the performance of COE alone (3895 individuals) with COE plus vital staining (4080 individuals) with biopsy and long-term follow-up through a National Cancer Registry (Su 2010). This study found a higher detection rate for oral cavity cancer in the conventional oral examination plus vital rinsing adjunct trial arm.The detection rate of oral pre-malignant lesions and malignant lesions after referral was 4.6% in the conventional oral examination plus vital rinsing arm; 4.4% in conventional oral examination alone. This resulted in a ratio of 1.05 (95% CI 0.74 to 1.41); an incidence rate of oral cancer ($x10^{-5}$) of 28 compared to 35.4 and relative incidence rate of 0.79 (95%CI 0.24 to 1.23). However, the initial screen positive rate was higher in the vital rinsing arm of the trial (9.5% and 8.3%).

When we consider the trial arms independently, the estimates of sensitivity and specificity for the target condition of oral cancer in the trial arm of COE alone were 0.50 (95% CI 0.12 to 0.88) and 0.92 (0.91 to 0.93) with a prevalence of 0.15%; the corresponding values for the COE with vital rinsing adjunct were 0.40 (95% CI 0.05 to 0.85) and 0.91 (0.90 to 0.91) with a prevalence of 0.13%.

A summary is given in the Summary of findings 3.

DISCUSSION

Summary of main results

Thirteen studies were identified for inclusion evaluating the diagnostic accuracy of conventional oral examination (COE), vital rinsing and mouth self examination. The studies were diverse in nature with substantial variations in sample prognostic risk factors, nature of screening test, the experience of personnel conducting the index test, verification of screen negative and screen positive individuals, exclusion of individuals from the analysis and large variation in incidence of disease (including register-based studies) across included studies. Consequently, the decision was taken that a meta-analysis of the included studies by index test was inappropriate. This is in contrast to some previously published systematic reviews (e.g. Downer 2004; Moles 2002).

Taken as a body of evidence, the overall quality of the studies was variable both within and between index tests with only one study (Julien 1995a) of COE being judged as overall low risk of bias and overall low concern regarding applicability (Figure 1). Many of the studies did not fully report on the characteristics and risk factors of the study sample, particularly important when assessing the applicability of the results. In five studies the participants could be considered as 'high risk' individuals and consequently their findings are of concern to the applicability of the review question.

Prevalence of potentially malignant disorders (PMD) or malignancy in the diagnostic test accuracy study samples ranged from 1.4% to 50.9% over the different index tests. Estimates should be interpreted with respect to the diagnostic test accuracy study prevalence levels. A low prevalence of the target condition effectively results in a lower sample size for diseased participants and for the calculation of sensitivity. For COE, sensitivity estimates were highly variable for study level prevalence analogous to those in the population, and ranged from 0.50 (95% confidence interval (CI) 0.07 to 0.93) to 0.99 (95% CI 0.97 to 1.00). The lower specificity values observed in the two studies where prevalence was significantly higher than would normally be observed (20% and 50%) the comparably lower specificity estimates can be explained at least in part by the higher prevalence. The variation in prevalence is reflective of the flow and timing of participants through the studies, particularly the process of investigation which was quite different from the flow and timing of the remaining included studies. All screened positive participants were offered the reference standard and all participants who attended the referral centre for subsequent verification received the reference standard. A random sample of participants screened negative received differential verification by the project dentist (diagnostic test accuracy evaluation samples of 2193 screen positive and 1350 screen negative (Warnakulasuriya 1991) and 660 screen positive 1212 and screen negative (Warnakulasuriya 1990)). For the two studies of mouth self examination, sensitivity values were 0.18 (95% CI 0.13 to 0.24) and 0.33 (95% CI 0.10 to 0.65) for mouth self examination. The one study that directly compared COE with COE plus vital rinsing in a randomised controlled trial found a higher detection rate for PMD in the trial arm with the vital rinsing adjunct.

Index tests at a prevalence reported in the population (between 1% and 5%) were better at correctly classifying the absence of PMD or oral cavity cancer in disease-free individuals than classifying the presence in diseased individuals. A false negative result from a screening programme would mean that the individuals with PMD or oral cavity cancer would not be referred for further investigations; a false positive result would mean a number of individuals without PMD or oral cavity cancer would receive a positive screening result, possibly resulting in further excisional investigations for the patient. Whereas the false positive results could and would no doubt have financial and other resource implications following inappropriate referral, the false negative results indicate that people with PMD or oral cavity cancer will be missed, possibly to be diagnosed at a later date when the disease will be more severe. For mouth self examination, the evidence is equivocal, with poor values of both sensitivity and specificity in one study. In the other study, a high value of specificity was accompanied by a very low sensitivity value. The prevalence of PMD or oral cavity cancer was high (10.6% and 22.6%) in both studies.

Strengths and weaknesses of the review

The utility of this review is limited in part by the number of included studies. A small number of potentially eligible studies were excluded on the basis that the screened negative individuals did not receive or report a reference standard. As a result, the number of false negatives could not be determined. Primary studies of more recently developed index tests were case-control studies and consequently ineligible for inclusion through study design. We took the decision to exclude case-control studies at the protocol stage owing to the potential for over estimation of diagnostic accuracy with this design. However, this has meant that the index tests evaluated in this review do not include those based on newer technologies. We would anticipate that those index tests showing promise at this present time, would be further evaluated with a more robust study design and therefore be eligible for inclusion in updates of this review.

Following on from previous systematic reviews in this area (e.g. Downer 2004), a further five diagnostic accuracy studies have been identified and were eligible for inclusion in this review. The

main strength of this review is that it evaluated the diagnostic accuracy of conventional oral examination, vital rinsing and mouth self examination. All included studies were assessed for methodological quality using the QUADAS-2 tool which we specifically adapted for this review. This enabled the quality of the evidence to be considered in conjunction with the diagnostic estimates.

Due to the substantial diversity in the nature of the included studies and the characteristics of the participants it was not appropriate to pool the data. Whilst this is not a weakness of the review, the failure to provide summary estimates of sensitivity and specificity, in contrast to previous systematic reviews, could be regarded as a limitation. The range of sensitivity values is likely to have been influenced by the considerable heterogeneity across the studies. In future updates should more homogeneous studies be included in the review, it would be informative to evaluate the influence of risk factors on estimates of diagnostic accuracy. However, we acknowledge that there was a lack of reported detail in a number of the included studies regarding the presence or absence of important risk factors such as smoking, betel quid chewing and alcohol consumption.

Participants were recruited into studies that had used a wide range of criteria from opportunistic screening programmes in company headquarters to mass screening programmes in South East Asia. The World Health Organization defines screening as "the application of a test or tests to people who are apparently free from the disease in question in order to distinguish between those that have the disease from those who probably do not" (Wilson 1968). A difficulty with a number of the included studies was determining how representative the screened population were given the settings for recruitment such as: a company's headquarters, hospital out-patient departments and tertiary treatment centres. It could be argued that the latter sample represents a distinct population with a much higher risk of developing new disease due to field change and one where clinicians are likely to have a higher index of suspicion. Prevalence of the included studies was in line with what would be expected; Napier 2008 argues that most authorities agree that this lies between 1% and 5%. However, the sample prevalence was particularly high in two studies of COE (Mathew 1997 10.3%, Ikeda 1995 9.7%) and one study of mouth self examination (Scott 2010 22.6%). In two studies of COE (Warnakulasuriya 1990; Warnakulasuriya 1991) the sample prevalence calculated from the two by two table evaluating the diagnostic test accuracy was particularly high at 21.6% and 50.9%. The screen positive prevalence for these studies was more in line with population prevalence at 2.25% and 6.23%.

The definition of a positive lesion was relatively consistent across all the studies, although in some studies (e.g. Mehta 1986), a positive screen could include 'growths suggestive of oral cancer' or referable lesions that were neither oral cavity cancer or PMD. Similarly, the definition of the target condition in the index test differed from that in the reference standard in some studies. In another study there was a lack of consistent definition and use of the target condition for the index and reference tests. As a potential source of bias, it was not always clear whether the reference standard had been interpreted with or without knowledge of the index test.

The use of cancer registries or other registries as a reference standard (e.g. Chang 2011; Su 2010) can be methodologically problematic, particularly if there is a mismatch in the target condition being evaluated and the outcome documented in the registry. For example cancer registries are unlikely to hold data on PMDs that have not undergone malignant transformation, inducing a mismatch in the target condition being detected by the index test and the outcome recorded in the registry. Differential verification bias can occur if screened positive participants receive biopsy as a reference standard whilst the screened negative participants are assessed through a national cancer registry alone. If there is potential for malignant transformation within the duration of follow-up then follow-up through registry could be appropriate. Careful thought should be given to the target condition of the index and reference standard and whether this information will be adequately recorded in the registry.

Applicability of findings to the review question

Concerns regarding applicability arose from targeted patient selection of high risk groups for the patient selection domain, where participants in five of the 13 had either a previous history of head and neck cancer or were older tobacco smokers. For example, participants in one study conducted in a tertiary care clinic (Chang 2011) were all males; and another study recruited former head and neck cancer patients undergoing routine surveillance visits (Sweeny 2011). Studies with unclear concerns over in this domain were those that had omitted important information on patient or study characteristics which meant that we were unable to determine whether the participants and settings matched the review question. There was low concern regarding applicability for the index test domain for all studies. An unclear judgement for applicability for the reference standard was given to one study where six people had been identified from the target population to act as the reference standard (Elango 2011). Although exposed to training, it is questionable whether trained lay people could act as a reference standard and there was some concern that the index test and reference test may have been conducted simultaneously for those who had not responded initially. A second study (Sweeny 2011) was also judged to be at unclear applicability on this domain. There was low concern regarding applicability for the remaining 11 studies.

AUTHORS' CONCLUSIONS

Implications for practice

There are known clinical and methodological difficulties associated with screening for potentially malignant disorders (PMDS) and cancer of the lip or oral cavity. These include the relatively low incidence rates, the reluctance of screened positive individuals to attend for follow-up, a lack of linear transition between pre-malignant and malignant states (Reibel 2003), disagreement over disease management (Warnakulasuriya 2009) and the relative cost-effectiveness of mass, selective and opportunistic screening programmes (Brocklehurst 2011).

A recent systematic review examined whether screening programmes for oral cancer could detect the disease and reduce the associated mortality of the condition (Brocklehurst 2013). One cluster randomised clinical trial was identified from Kerala in India. The screening programme comprised of four cycles over a 15-year period and involved 13 clusters with 191,873 participants. There was no statistically significant difference in the oral cancer mortality rates between the screened group (15.4/100,000 person-years) and

the control group (17.1/100,000 person-years). However, when only high risk individuals were included in the analysis (users of tobacco or alcohol or both), there was a reported reduction of 24% in the mortality rate. A statistically significant reduction was also found in the number of individuals diagnosed with late stage disease in the screened group (risk ratio 0.81; 95% confidence interval 0.70 to 0.93). No harms were reported but the study was assessed to be of high risk of bias. Across the four cycles (15 years) of the programme, the reported sensitivity of the visual examination in detecting oral cancer was 67.4% (188/279). No information on the specificity or the positive predictive value of the programme was recorded. However, the latter was calculated based on the published data from the study as the number of screen-selected oral cancers as a proportion of total screened positive subjects (confirmed by biopsy), which was 86.5% for oral cancer.The cost-effectiveness of this study was considered to meet the standards of the World Health Organization (Subramanian 2009). Selective screening of high risk groups and opportunistic screening may reduce costs (Speight 2006), but many high risk patients do not attend general dental practices (Netuveli 2006).

The lack of any formal registration for PMD, in contrast to malignancy, makes it difficult to estimate possible reductions in mortality due to a screening programme aimed at precursor lesions. In addition, the efficacy of the early management of PMDs is a controversial area and the evidence base has recently been challenged (Holmstrup 2007; Holmstrup 2009). Holmstrup has demonstrated that even if lesions are surgically removed, the risk of malignant change may remain since the lesion represents only a small area in a field of damaged mucosa, any part of which may progress to malignancy.

The results of this review suggest that using the conventional oral examination (COE) for screening for PMD and oral cancer has a variable degree of sensitivity (greater than 0.70 in six of the 10 studies) and a consistently high value for specificity (greater than 0.90 in all eight studies). However, there was considerable clinical heterogeneity in the participants forming the study samples, the application of the index test and reference standard and the flow and timing of the process. Exploring the primary studies for sources of heterogeneity has not shown any single factor to consistently influence the accuracy of the screening test. In terms of test accuracy, there is limited evidence of performance in each of the different settings, with clinicians or non-clinicians carrying out the index test etc. which means that the current evidence base is limited, though COE has been shown to have good estimates of both sensitivity and specificity in some studies. Further, even though the evidence of accuracy is not consistently strong, there is some evidence (Brocklehurst 2013) that implementing COE as a component of a population screening programme can reduce mortality and produce stage-shift in a high risk population. Should such findings be replicated in other studies then it could be argued that explicit testing of the test accuracy per se of the COE is unnecessary, given the positive outcomes on mortality. Emphasis could be placed on the effectiveness of screening programmes, of which COE is a component, in reducing morbidity and mortality. This should be supplemented with information on the consequences of false negative and false positive screens.

There is insufficient evidence to deviate from the conclusions of the American Dental Association that oral cancer screening may detect PMDs and cancer of the lip or oral cavity (Rethman 2010). General dental practitioners and dental care professionals should remain vigilant for signs of PMD and oral cancer whilst performing routine oral examinations in practice.

The sensitivity estimates for mouth self examination were lower than for COE, though these studies were on different participant samples and should not be directly compared. There is insufficient evidence to satisfactorily determine the diagnostic test accuracy of mouth self examination as part of an organised screening programme.

Implications for research

It is clear that there are some methodological shortcomings in the studies included in this review. The QUADAS-2 tool has provided a robust means of assessing the methodological quality of the included studies. There is now an opportunity to use this framework, along with the guidance from the Cochrane Diagnostic Test Accuracy Editorial Group, to ensure that future studies are conducted in a robust manner, with particular attention paid to the design of the study in the four domains of the QUADAS-2 tool. It is imperative that studies are reported with sufficient information to allow judgement of the merits of the study and its applicability to the review being undertaken. Reporting according to the STARD checklist should facilitate this process. In particular, results have been promising in the workplace setting, and for some opportunistic screening studies.

The population and participant selection should be clearly stated and carried out to reduce the possibility of sampling bias, preferably using a consecutive sample. The index test should be undertaken by trained and calibrated screeners, whose threshold for agreement should be stated a priori. The reference standard should be both accurate and pragmatic to account for the practical considerations involved in establishing the initial diagnostic test accuracy component of large population screening programmes. For such programmes it is not necessary to apply the reference standard to the entire programme's participants, rather an initial evaluation of test accuracy should be established on a sizeable number of participants prior to commencement of the screening programme proper. It is also important to utilize reference standards that capture all the target conditions under question, not just those that are likely to be identified through cancer registries. Finally, the flow and timing of the diagnostic test accuracy study should ensure that the reference standard is undertaken within a short time frame after the index test, given the potential for PMD to undergo malignant transformation and for it to be applied after the index test to avoid bias being introduced. Where long-term follow-up is used as a reference standard, measures should be taken to minimise attrition. Further research on ways to maximise initial participation rates and also follow-up rates for those screened positive is warranted.

ACKNOWLEDGEMENTS

We would like to thank NHS Education Scotland and the Scottish Dental Clinical Effectiveness Programme for their support on this review; Anne Littlewood (Cochrane Oral Health Group) for her advice on searching the literature; Derek Richards for his thoughtful feedback on the protocol; Richard Macey, Kathryn Taylor and Joanne Leese for help with data extraction and Luisa Fernandez Mauleffinch and Philip Riley (Cochrane Oral Health Group) for

their assistance in facilitating this review; Jan Clarkson and Helen Worthington for initiating and facilitating this review.

REFERENCES

References to studies included in this review

Chang 2011 *{published data only}*

Chang IH, Jiang RS, Wong YK, Wu SH, Chen FJ, Liu SA. Visual screening of oral cavity cancer in a male population: experience from a medical center. *Journal of the Chinese Medical Association* 2011;**74**(12):561-6.

Downer 1995 *{published data only}*

Downer MC, Evans AW, Hughes Hallett CM, Jullien JA, Speight PM, Zakrzewska JM. Evaluation of screening for oral cancer and precancer in a company headquarters. *Community Dentistry and Oral Epidemiology* 1995;**23**(2):84-8.

Elango 2011 *{published data only}*

Elango KJ, Anandkrishnan N, Suresh A, Iyer SK, Ramaiyer SK, Kuriakose MA. Mouth self-examination to improve oral cancer awareness and early detection in a high-risk population. *Oral Oncology* 2011;**47**(7):620-4.

Ikeda 1995 *{published data only}*

Ikeda N, Downer MC, Ishii T, Fukano H, Nagao T, Inoue K. Annual screening for oral cancer and precancer by invitation to 60-year-old residents of a city in Japan. *Community Dental Health* 1995;**12**(3):133-7.

Julien 1995 *{published data only}*

Jullien JA, Downer MC, Zakrzewska JM, Speight PM. Evaluation of a screening test for the early detection of oral cancer and precancer. *Community Dental Health* 1995;**12**:3-7.

Julien 1995a *{published data only}*

Jullien JA, Downer MC, Zakrzewska JM, Speight PM. Evaluation of a screening test for the early detection of oral cancer and precancer. *Community Dental Health* 1995;**12**:3-7.

Mathew 1997 *{published data only}*

* Mathew B, Sankaranarayanan R, Sunilkumar KB, Kuruvila B, Pisani P, Nair MK. Reproducibility and validity of oral visual inspection by trained health workers in the detection of oral precancer and cancer. *British Journal of Cancer* 1997;**76**(3):390-4.

Ramadas K, Sankaranarayanan R, Jacob BJ, Thomas G, Somanathan T, Mahé C, et al. Interim results from a cluster randomized controlled oral cancer screening trial in Kerala, India. *Oral Oncology* 2003;**39**(6):580-8.

Sankaranarayanan R, Mathew B, Jacob BJ, Thomas G, Somanathan T, Pisani P, et al. Early findings from a community-based, cluster-randomized, controlled oral cancer screening trial in Kerala, India. The Trivandrum Oral Cancer Screening Study Group. *Cancer* 2000;**88**(3):664-73.

Sankaranarayanan R, Ramadas K, Thara S, Muwonge R, Thomas G, Anju G, et al. Long term effect of visual screening on oral cancer incidence and mortality in a randomized trial in Kerala, India. *Oral Oncology* 2013;**49**(4):314-21.

Mehta 1986 *{published data only}*

Mehta FS, Gupta PC, Bhonsle RB, Murti PR, Daftary DK, Pindborg JJ. Detection of oral cancer using basic health workers in an area of high oral cancer incidence in India. *Cancer Detection and Prevention* 1986;**9**(3-4):219-25.

Scott 2010 *{published data only}*

Scott SE, Rizvi K, Grunfeld EA, McGurk M. Pilot study to estimate the accuracy of mouth self-examination in an at-risk group. *Head and Neck* 2010;**32**(10):1393-401.

Su 2010 *{published data only}*

Su WW, Yen AM, Chiu SY, Chen TH. A community-based RCT for oral cancer screening with toluidine blue. *Journal of Dental Research* 2010;**89**(9):933-7.

Sweeny 2011 *{published data only}*

Sweeny L, Dean NR, Magnuson JS, Carroll WR, Clemons L, Rosenthal EL. Assessment of tissue autofluorescence and reflectance for oral cavity cancer screening. *Journal of Otolaryntology - Head and Neck Surgery* 2011;**145**(6):956-60.

Warnakulasuriya 1990 *{published data only}*

Warnakulasuriya KA, Ekanayake AN, Sivayoham S, Stjernswärd J, Pindborg JJ, Sobin IH, et al. Utilization of primary health care workers for early detection of oral cancer and precancer cases in Sri Lanka. *Bulletin of the World Health Organization* 1984;**62**(2):243-50.

* Warnakulasuriya S, Pindborg JJ. Reliability of oral precancer screening by primary health care workers in Sri Lanka. *Community Dental Health* 1990;**7**(1):73-9.

Warnakulasuriya 1991 *{published data only}*

Warnakulasuriya KA, Nanayakkara BG. Reproducibility of an oral cancer and precancer detection program using a primary health care model in Sri Lanka. *Cancer Detection and Prevention* 1991;**15**(5):331-4.

References to studies excluded from this review

Bhalang 2008 *{published data only}*

Bhalang KA, Suesuwan A, Dhanuthai K, Sannikorn P, Luangjarmekorn L, Swasdison S. The application of acetic acid in the detection of oral squamous cell carcinoma. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontics* 2008;**106**(3):371-6.

Bowles 1973 *{published data only}*

Bowles DA Jr, Levin ES, Hasler JF. A comparison of two cytologic methods for mass screening against oral cancer. *Journal of Oral Surgery* 1973;**31**(11):830-6.

Chen 2007 *{published data only}*

Chen YW, Lin JS, Fong JH, Wang IK, Chou SJ, Wu CH, et al. Use of methylene blue as a diagnostic aid in early detection of oral cancer and precancerous lesions. *British Journal of Oral and Maxillofacial Surgery* 2007;**45**(7):590-1.

Csépe 2007 *{published data only}*

Csépe P, Bánóczy J, Dombi C, Forrai J, Gyenes M, Döbrossy L. Model program for screening oral cancers in the Roma population. *Magyar Onkologia* 2007;**51**(2):95-101.

Fernández Garrote 1995 *{published data only}*

Fernández Garrote L, Sankaranarayanan R, Lence Anta JJ, Rodriguez Salvá A, Maxwell Parkin D. An evaluation of the oral cancer control program in Cuba. *Epidemiology* 1995;**6**(4):428-31.

Hapner 2011 *{published data only}*

Hapner ER, Wise JC. Results of a large-scale head and neck cancer screening of an at-risk population. *Journal of Voice* 2011;**25**(4):480-3.

Huber 2004 *{published data only}*

Huber MA, Bsoul SA, Terzhalmy GT. Acetic acid wash and chemiluminescent illumination as an adjunct to conventional oral soft tissue examination for the detection of dysplasia: a pilot study. *Quintessence International* 2004;**35**(5):378-84.

Huff 2009 *{published data only}*

Huff K, Stark PC, Solomon LW. Sensitivity of direct tissue fluorescence visualization in screening for oral premalignant lesions in general practice. *General Dentistry* 2009;**57**(1):34-8.

Leocata 2007 *{published data only}*

Leocata P, D'Alessandro A, D'Amario M, Barone A, Giannoni M. A screening program for the early detection and prevention of oral cancer. *Journal of Plastic Dermatology* 2007;**3**(1):71-5.

Lim 2003 *{published data only}*

Lim K, Moles DR, Downer MC, Speight PM. Opportunistic screening for oral cancer and precancer in general dental practice: results of a demonstration study. *British Dental Journal* 2003;**194**(9):497-502.

Nagao 2000 *{published data only}*

* Nagao T, Ikeda N, Fukano H, Miyazeki H, Yano M, Warnakulasuriya S. Outcome following a population screening programme for oral cancer and precancer in Japan. *Oral Oncology* 2000;**36**:340-6.

Nagao T, Warnakulasuriya S. Annual screening for oral cancer detection. *Cancer Detection and Prevention* 2003;**27**:333-7.

Nagao T, Warnakulasuriya S, Ikeda N, Fukano H, Fujiwara K, Miyazaki H. Oral cancer screening as an integral part of general health screening in Tokoname City, Japan. *Journal of Medical Screening* 2000;**7**(4):203-8.

Oh 2007 *{published data only}*

Oh ES, Laskin DM. Efficacy of the ViziLite system in the identification of oral lesions. *Journal of Oral and Maxillofacial Surgery* 2007;**65**(3):424-6.

Poh 2007 *{published data only}*

Poh CF, Hislop G, Currie B, Lee R, Sikorski S, Zed C, et al. Oral cancer screening in a high-risk underserved community--Vancouver Downtown Eastside. *Journal of Health Care for the Poor and Underserved* 2007;**18**(4):767-78.

Srivasta 1971 *{published data only}*

Srivastava YC, Mathur MN. Toluidine blue staining test for early detection of oral cancer: an adjunct to biopsy. *Dental Digest* 1971;**77**(7):400-3.

Vahidy 1972 *{published data only}*

Vahidy NA, Zaidi SH, Jafarey NA. Toludine blue test for detection of carcinoma of the oral cavity: an evaluation. *Journal of Surgical Oncology* 1972;**4**(5):434-8.

Warnakulasuriya 2010 *{published data only}*

Warnakulasuriya S, Kashyap R, Dasanayake AP. Is workplace screening for potentially malignant oral disorders feasible in India?. *Journal of Oral Pathology and Medicine* 2010;**39**(9):672-6.

References to studies awaiting assessment

Barrellier 1980 *{published data only}*

Barrellier P, Rame JP, Chasle J, Souquières Y. Detection of buccal cavity cancers with toluidine blue (author's transl). *Revue de Stomatologie et de Chirurgie Maxillo-Faciale* 1980;**81**(6):364-7.

Barrellier 1982 *{published data only}*

Barrellier P, Rame JP, Chasle J, Souquières Y, Lecacheux B. Detection of cancer of the oral cavity using toluidine blue. *Actualités Odonto-Stomatologiques* 1982;**36**(137):87-92.

Marzouki 2012 *{published data only}*

Marzouki HZ, Tuong Vi Vu T, Ywakim R, Chauvin P, Hanley J, Kost KM. Use of fluorescent light in detecting malignant and premalignant lesions in the oral cavity: a prospective, single-blind study. *Journal of Otolaryngology - Head and Neck Surgery* 2012;**41**(3):164-8.

Vacher 1999 *{published data only}*

Vacher C, Legens M, Rueff B, Lezy JP. Screening of cancerous and precancerous lesions of the oral mucosa in an at-risk population. *Revue de Stomatologie et de Chirurgie Maxillo-Faciale* 1999;**100**:180-3.

References to ongoing studies

Kulak 2010 *{published data only}*

Kulak J, Franzmann E. The use of autofluorescence in detection of oral lesions. *The Laryngoscope* 2010;**120**(S3):S40.

Additional references

American Cancer Society 1992

American Cancer Society. Update January 1992: the American Cancer Society guidelines for the cancer-related checkup. *CA: a Cancer Journal for Clinicians* 1992;**42**(1):44-5.

Bessell 2011

Bessell A, Glenny A-M, Furness S, Clarkson JE, Oliver R, Conway DI, et al. Interventions for the treatment of oral and oropharyngeal cancers: surgical treatment. *Cochrane Database of Systematic Reviews* 2011, Issue 9. [DOI: 10.1002/14651858.CD006205.pub3]

Brinkmann 2011

Brinkmann O, Kastratovic DA, Dimitrijevic MV, Konstantinovic VS, Jelovac DB, Antic J, et al. Oral squamous cell carcinoma detection by salivary biomarkers in a Serbian population. *Oral Oncology* 2011;**47**(1):51-5.

Brocklehurst 2010

Brocklehurst PR, Baker SR, Speight PM. Oral cancer screening: what have we learnt and what is there still to achieve?. *Future Oncology* 2010;**6**(2):299-304.

Brocklehurst 2011

Brocklehurst PR, Ashley JR, Tickle M. Patient assessment in general dental practice - risk assessment or clinical monitoring?. *British Dental Journal* 2011;**210**(8):351-4.

Brocklehurst 2013

Brocklehurst P, Kujan O, O'Malley LA, Ogden G, Shepherd S, Glenny AM. Screening programmes for the early detection and prevention of oral cancer. *Cochrane Database of Systematic Reviews* 2013, Issue 11. [DOI: 10.1002/14651858.CD004150.pub4]

Buchen 2011

Buchen L. Cancer: Missing the mark. Why is it so hard to find a test to predict cancer?. *Nature* 2011;**471**(7339):428-32.

Cancer Research UK

Cancer Research UK. http://www.cancerresearchuk.org/cancer-info/cancerstats/types/oral/uk-oral-cancer-statistics.

Conway 2008

Conway DI, Petticrew M, Marlborough H, Berthiller J, Hashibe M, Macpherson LM. Socioeconomic inequalities and oral cancer risk: a systematic review and meta-analysis of case-control studies. *International Journal of Cancer* 2008;**122**(12):2811-9.

Downer 2004

Downer MC, Moles DR, Palmer S, Speight PM. A systematic review of test performance in screening for oral cancer and precancer. *Oral Oncology* 2004;**40**(3):264-73.

Faggiano 1997

Faggiano F, Partanen T, Kogevinas M, Boffetta P. Socioeconomic differences in cancer incidence and mortality. In: Kogevinas M, Pearce N, Susser M, Boffetta P editor(s). Social Inequalities and Cancer. Lyon: IARC Scientific Publications No 138. International Agency for Research in Cancer, 1997.

Fedele 2009

Fedele S. Diagnostic aids in the screening of oral cancer. *Head & Neck Oncology* 2009;**1**:5.

Ferlay 2010

Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer* 2010;**127**(12):2893-917.

Freedman 2007

Freedman ND, Abnet CC, Leitzmann MF, Hollenbeck AR, Schatzkin A. Prospective investigation of the cigarette smoking-head and neck association by sex. *Cancer* 2007;**110**(7):1593-601.

Furness 2011

Furness S, Glenny AM, Worthington HV, Pavitt S, Oliver R, Clarkson JE, et al. Interventions for the treatment of oral cavity and oropharyngeal cancer: chemotherapy. *Cochrane Database of Systematic Reviews* 2011, Issue 4. [DOI: 10.1002/14651858.CD006386.pub3]

Garg 2005

Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 2005;**293**(10):1223-38.

Glenny 2010

Glenny AM, Furness S, Worthington HV, Conway DI, Oliver R, Clarkson JE, et al. Interventions for the treatment of oral cavity and oropharyngeal cancer: radiotherapy. *Cochrane Database of Systematic Reviews* 2010, Issue 12. [DOI: 10.1002/14651858.CD006387.pub2]

Holmstrup 2006

Holmstrup P, Vedtofte P, Reibel J, Stoltze K. Long-term treatment outcome of oral premalignant lesions. *Oral Oncology* 2006;**42**(5):461-74.

Holmstrup 2007

Holmstrup P, Vedtofte P, Reibel J, Stoltze K. Oral premalignant lesions: is a biopsy reliable?. *Journal of Oral Pathology and Medicine* 2007;**36**(5):262–6.

Holmstrup 2009

Holmstrup P. Can we prevent malignancy by treating premalignant lesions?. *Oral Oncology* 2009;**45**:549-50.

Jaber 2003

Jaber MA, Porter SR, Speight P, Eveson JW, Scully C. Oral epithelial dysplasia: clinical characteristics of western European residents. *Oral Oncology* 2003;**39**(6):589-96.

Jemal 2008

Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, et al. Cancer statistics, 2008. *CA: A Cancer Journal for Clinicians* 2008;**58**(2):71-96.

La Vecchia 1997

La Vecchia C, Tavani A, Franceschi S, Levi F, Corrao G, Negri E. Epidemiology and prevention of oral cancer. *Oral Oncology* 1997;**33**(5):302-12.

Landis 1977

Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**(1):159-74.

Lee 2009

Lee JM, Garon E, Wong DT. Salivary diagnostics. *Orthodontics and Craniofacial Research* 2009;**12**(3):206-11.

Leeflang 2008

Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM, Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. *Annals of Internal Medicine* 2008;**149**(12):889-97.

Leston 2010

Seoane Leston J, Diz Dios P. Diagnostic clinical aids in oral cancer. *Oral Oncology* 2010;**46**(6):418-22.

Li 2006

Li Y, Elashoff D, Oh M, Sinha U, St John MA, Zhou X, et al. Serum circulating human mRNA profiling and its utility for oral cancer detection. *Journal of Clinical Oncology* 2006;**24**(11):1754-60.

Lingen 2008

Lingen MW, Kalmar JR, Karrison T, Speight PM. Critical evaluation of diagnostic aids for the detection of oral cancer. *Oral Oncology* 2008;**44**(1):10-22.

Liu 2006

Liu JL, Wyatt JC, Deeks JJ, Clamp S, Keen J, Verde P, et al. Systematic reviews of clinical decision tools for acute abdominal pain. *Health Technology Assessment* 2006;**10**(47):1-167.

Liu 2012

Liu JLY, Walsh T, Kerr AR, Lingen M, Brocklehurst P, Ogden G, et al. Diagnostic tests for oral cancer and potentially malignant disorders in patients presenting with clinically evident lesions. *Cochrane Database of Systematic Reviews* 2012, Issue 12. [DOI: 10.1002/14651858.CD010276]

Lodi 2006

Lodi G, Sardella A, Bez C, Demarosi F, Carrassi A. Interventions for treating oral leukoplakia. *Cochrane Database of Systematic Reviews* 2006, Issue 4. [DOI: 10.1002/14651858.CD001829.pub3]

Macfarlane 1995

Macfarlane GJ, Zheng T, Marshall JR, Boffetta P, Niu S, Brasure J, et al. Alcohol, tobacco, diet and the risk of oral cancer: a pooled analysis of three case-control studies. *European Journal of Cancer. Part B, Oral Oncology* 1995;**31B**(3):181-7.

Mashberg 1988

Mashberg A, Feldman LJ. Clinical criteria for identifying early oral and oropharyngeal carcinoma: erythroplasia revisited. *American Journal of Surgery* 1988;**156**(4):273-5.

Mashberg 1995

Mashberg A, Samit A. Early diagnosis of asymptomatic oral and oropharyngeal squamous cancers. *CA: A Cancer Journal for Clinicians* 1995;**45**(6):328-51.

McGurk 2010

McGurk M, Scott SE. The reality of identifying early oral cancer in the general dental practice. *British Dental Journal* 2010;**208**(8):347-51.

Moles 2002

Moles DR, Downer MC, Speight PM. Meta-analysis of measures of performance reported in oral cancer and precancer screening studies. *British Dental Journal* 2002;**192**(6):340-4.

Napier 2008

Napier SS, Speight PM. Natural history of potentially malignant oral lesions and conditions: an overview of the literature. *Journal of Oral Pathology and Medicine* 2008;**37**(1):1-10.

Netuveli 2006

Netuveli G, Sheiham A, Watt RG. Does the 'inverse screening law' apply to oral cancer screening and regular dental check-ups?. *Journal of Medical Screening* 2006;**13**(1):47-50.

Ogden 2005

Ogden GR. Alcohol and oral cancer. *Alcohol* 2005;**35**(3):169-73.

Parkin 2001

Parkin DM. Global cancer statistics in the year 2000. *Lancet Oncology* 2001;**2**(9):533-43.

Patton 2008

Patton LL, Epstein JB, Kerr AR. Adjunctive techniques for oral cancer examination and lesion diagnosis: a systematic review of the literature. *Journal of American Dental Association* 2008;**139**(7):896-905.

Petti 2003

Petti S. Pooled estimate of world leukoplakia prevalence: a systematic review. *Oral Oncology* 2003;**39**(8):770-80.

Reibel 2003

Reibel J. Prognosis of oral pre-malignant lesions: significance of clinical, histopathological, and molecular biological characteristics. *Critical Reviews in Oral Biology and Medicine* 2003;**14**(1):47-62.

Rethman 2010

Rethman MP, Carpenter W, Cohen EE, Epstein J, Evans CA, Flaitz CM, et al. Evidence-based clinical recommendations regarding screening for oral squamous cell carcinomas. *Journal of the American Dental Association* 2010;**141**(5):509-20.

RevMan 2012 [Computer program]

The Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan). Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2012.

Rogers 2009

Rogers SN, Brown JS, Woolgara JA, Lowe D, Magennis P, Shaw RJ, et al. Survival following primary surgery for oral cancer. *Oral Oncology* 2009;**45**(3):201-11.

Rosin 2006

National Institute of Dental and Craniofacial Research. Bringing the promise of molecular medicine to oral cancer screening. http://www.nidcr.nih.gov/Research/ResearchResults/InterviewsOHR/TIS032006.htm (accessed 18th April 2011).

Rusthoven 2010

Rusthoven KE, Raben D, Song JI, Kane M, Altoos TA, Chen C. Survival and patterns of relapse in patients with oral tongue cancer. *Journal of Oral and Maxillofacial Surgery* 2010;**68**(3):584-9.

Sankaranarayanan 1997

Sankaranarayanan R. Health care auxiliaries in the detection and prevention of oral cancer. *Oral Oncology* 1997;**33**(3):149-54.

Scully 2000b

Scully C, Shotts R. ABC of oral health. Mouth ulcers and other causes of orofacial soreness and pain. *BMJ* 2000;**321**(7254):162-5.

Scully 2009

Scully C, Bagan J. Oral squamous cell carcinoma overview. *Oral Oncology* 2009;**45**(4-5):301-8.

Silverman 1984

Silverman S Jr, Gorsky M, Lozada F. Oral leukoplakia and malignant transformation. A follow-up study of 257 patients. *Cancer* 1984;**53**(3):563-8.

Speight 2006

Speight PM, Palmer S, Moles DR, Downer MC, Smith DH, Henriksson M, et al. The cost-effectiveness of screening for oral cancer in primary care. *Health Technology Assessment* 2006;**10**(14):1-144.

Subramanian 2009

Subramanian S, Sankaranarayanan R, Bapat B, Somanathan T, Thomas G, Matthew B, et al. Cost-effectiveness of oral cancer screening: results from a cluster randomized controlled trial in India. *Bulletin of the World Health Organization* 2009;**87**(3):200–6.

US Preventive Services Task Force

US Preventive Services Task Force. http://www.uspreventiveservicestaskforce.org/ (accessed 8th April 2011).

van der Waal 2009

van der Waal I. Potentially malignant disorders of the oral and oropharyngeal mucosa: terminology, classification and present concepts of management. *Oral Oncology* 2009;**45**(4-5):317-23.

Warnakulasuriya 2007

Warnakulasuriya S, Johnson NW, van der Waal I. Nomenclature and classification of potentially malignant disorders of the oral mucosa. *Journal of Oral Pathology and Medicine* 2007;**36**(10):575-80.

Warnakulasuriya 2009

Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. *Oral Oncology* 2009;**45**(4-5):309-16.

Whiting 2004

Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Annals of Internal Medicine* 2004;**140**(3):189-202.

Whiting 2011

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529-36.

Wilson 1968

Wilson JM, Jungner YG. Principles and practice of mass screening for disease. *Boletín de la Oficina Sanitaria Panamericana* 1968;**65**(4):281-393.

Wyatt 1995

Wyatt JC, Altman DG. Commentary: Prognostic models: clinically useful or quickly forgotten?. *BMJ* 1995;**311**:1539.

Yusof 2006

Yusof ZY, Netuveli G, Ramli AS, Sheiham A. Is opportunistic oral cancer screening by dentists feasible? An analysis of the patterns of dental attendance of a nationally representative sample over 10 years. *Oral Health and Preventive Dentistry* 2006;**4**(3):165-71.

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Chang 2011

Study characteristics	
Patient sampling	Method of patient selection: Optional screening programme at a tertiary referral centre in central Taiwan. "All male patients who visited our clinic (Otolaryngology or Dental Department) aged 18 or older were eligible for enrolment in this study." "Those who were reluctant to undergo oral screening were excluded"

Chang 2011 *(Continued)*

Patient characteristics and setting	13,878 patients enrolled from March 2005 to January 2010 Age: Mean age 54.6 years (sd 18.4 range 18 to 97 years) Sex: Male population, reasons for single sex sample not stated SES: Not stated Ethnicity: Not stated Stated risk factors: 2844 habitual smokers; 943 habitual betel quid chewers; 1955 habitual drinkers Previous history: Not stated Location: Taiwan Clinical setting: Tertiary academic medical centre. Veterans General Hospital
Index tests	Index test: "..visual screening of the oral cavity was performed by experienced otolaryngologists or dentists under adequate lighting and with proper instruments" Description of positive case definition by index test as reported: "A non-healing ulcer for more than 2 weeks, a persistent white or red lesion, a lesion that bled easily, or an irregular surface lesion inside the oral cavity were regarded as positive findings." Positive lesions indicative of oral cavity cancer Sequence of tests: Index followed by reference Training or calibration: Not stated Blinding of examiners: Not stated Conflict of interests: Authors declare no conflict of interest
Target condition and reference standard(s)	Target condition: Oral cavity cancer Reference standard: Punch biopsy with histopathology of abnormal lesions. "If the patient did not agree to further biopsy, follow-up was strongly recommended." Follow-up of entire cohort. "We further crosslinked the entire screened cohort with the Taiwan Cancer Registry database" Description of positive case definition by reference test as reported: Oral cavity cancer Training or calibration: Not stated Blinding of examiners: Not stated Prevalence of the target condition on the sample: 285/13,606 2.1%
Flow and timing	Time interval and any interventions between index test(s) and reference standard: Not reported Characteristics and proportion of individuals who did not receive the index test(s) and/or reference standard or excluded from analysis: "A total of 272 participants (37.5%) with abnormal oral cavity lesions were lost to follow-up and no further pathological report could be obtained." "In order not to confound further analyses, we excluded those with positive lesions/yet no further biopsy during the follow-up period. Although 272 participants were excluded from the final analysis, there was little impact on the power of the statistic analysis due to the large population size" Characteristics and proportion of individuals who received a reference standard other than examination and clinical evaluation by a specialist physician: "We further cross linked the entire screened cohort with the Taiwan Cancer Registry database." Not reported when this was done (follow-up time) for entire cohort
Comparative	

Chang 2011 *(Continued)*

Notes	Sensitivity and specificity data reported for oral cavity cancer. Index test target condition clinically suspicious oral lesions; reference standard target condition oral cancer.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was conflict of interest avoided?	Yes		
Where multiple index tests were used, were the results of the second index test interpreted without knowledge of the results of the first index test?			
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index	Unclear		

Chang 2011 *(Continued)*

test and reference standard?	
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	No
	High

Downer 1995

Study characteristics	
Patient sampling	Method of patient selection: Employees (40 years or over) in a workplace setting responding to a screening invitation. Screening programme was widely publicised through in-house magazine, information leaflets, video in hallway. Participation rate 53%
Patient characteristics and setting	292/553 (53%) of workers responded to the screening invitation. Additional 17 screened from a separate site Age: ≥40 years Sex: Not stated SES: 31.8% lower occupational level, 68.2% management grade or above Ethnicity: Not stated Stated risk factors: HPV - not stated; smoking - smokers included in sample but proportions not specified; alcohol - drinkers included in sample but proportions not specified Previous history: Not stated Location: Commercial company. London, UK Clinical setting: Onsite company dental practice
Index tests	Index test: Systematic visual examination by 2 general dental practitioners Description of positive case definition by index test as reported: "...if a white patch, red patch or ulcer of greater than two weeks duration was detected." Further qualified into lesions that should be regarded as malignant or pre-malignant (positive) and those to be regarded as negative Sequence of tests: Index test followed by reference standard Training or calibration: "...who had not received any specific training except for instruction in the screening procedure and the criteria for a positive or negative test." No specific training and standardisation of screeners nor calibration Blinding of examiners: Index test completed before reference standard Conflict of interests: Not stated
Target condition and reference standard(s)	Target condition: As for the index test: Carcinoma, leukoplakia, erythroplakia, lichen planus, lupus erythematosus, submucous fibrosis, actinic keratosis Reference standard: Visual examination by an oral medicine specialist with "...access to any relevant diagnostic aids, including biopsy if considered necessary"

Downer 1995 *(Continued)*

	Description of positive case definition by reference test as reported: "...if a white patch, red patch or ulcer of greater than two weeks duration was detected." Further qualified into lesions that should be regarded as malignant or pre-malignant (positive) and those to be regarded as negative Training or calibration: Not stated Blinding of examiners: Index test completed before reference standard. "...who was unaware of the findings of the screener" Prevalence of the target condition on the sample: 17/309 5.5%
Flow and timing	Time interval and any interventions between index test(s) and reference standard: Immediately following attendance at screening session: "After screening..." Characteristics and proportion of individuals who did not receive the index test(s) and/or reference standard or excluded from analysis: "A number of staff who were screened will not have been included in the evaluation since they were unable to attend at one of the dedicated sessions and were therefore not examined by the specialist diagnostician." Separate values for those attending the screening and reference standard examination not reported Characteristics and proportion of individuals who received a reference standard other than examination and clinical evaluation by a specialist physician: None
Comparative	
Notes	68.2% proportion of participants at management grade or above. 53% participation rate

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was conflict of interest avoided?	Yes		
Where multiple index tests were used, were the results of the second index test interpreted without			

Downer 1995 *(Continued)*

knowledge of the results of the first index test?			
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
		Low	

Elango 2011

Study characteristics	
Patient sampling	Method of patient selection: "The study population was distributed in two Panchayats (local administrative unit in villages) with 33 subunits. Brochures were sequentially distributed to all the houses in the subunits." After a lapse of 4 weeks "Health workers attempted to locate individuals up to a maximum of three times, incase they were unavailable during the first visit"
Patient characteristics and setting	Results available for 34,766/48,080 eligible participants. "48,080 (83.3%) subjects, above the age of 10 years, were eligible for the study" Age: Median age band 30-39 years Sex: 17,158 male 17,608 female SES: Not stated Ethnicity: Not stated Stated risk factors: Tobacco smoking / chewing pan 10,644; alcohol consumption 3844 Previous history: Not stated

Elango 2011 *(Continued)*

	Location: Kerala, India. "It was carried out in a high-risk population of 57,704, in the coastal villages of Kerala, India, where there is a high incidence of oral cancer and prevalence of risk factors" Clinical setting: Participants' own homes
Index tests	Index test: Mouth self examination in accordance with brochures specifically designed for this population. "A brochure was developed, which contained information on oral cancer, its risk factors and the methods to perform MSE. It also had instructions to report to the oral cancer-screening clinic, in case of identification of any suspicious lesions" Description of positive case definition by index test as reported: White patch, red patch, non-healing ulcers, difficulty in opening mouth, other oral symptoms (burning sensation) Sequence of tests: Index test followed by reference standard Training or calibration: Dedicated brochure instructed on mouth self examination technique Blinding of examiners: No description of timing or recording of mouth self examination in relation to visit by health worker 4 weeks after screening exam (mouth self examination could have been carried out concurrently) Conflict of interests: None. "The project was supported by Government of India, Department of Science and Technology, research grant (SSD/SCP/060/2005)"
Target condition and reference standard(s)	Target condition: Oral cancer and potentially malignant lesions Reference standard: "Six health workers recruited from the population wherein the study was conducted..." Description of positive case definition by reference test as reported: "The presence (including site and provisional diagnosis) and absence of potentially malignant oral lesions (ulcers, white or red patches, or lumps/swellings) were noted on a proforma" Training or calibration: "Six health workers underwent one month training on oral cancer in a comprehensive cancer center, which coordinated the study. The training consisted of a didactic course on oral cancer, its risk factors, clinical findings of potentially malignant and malignant oral lesions, and methods to perform oral visual examination. WHO Guide to epidemiology and diagnosis of oral mucosal diseases and conditions was used as the reference manual. The competence of the health workers was confirmed by a trained oral cancer specialist." Calibration not stated Blinding of examiners: Not stated Prevalence of the target condition on the sample: 219/34,766 0.63%
Flow and timing	Time interval and any interventions between index test(s) and reference standard: "After a lapse of 4 weeks, the trained health workers performed oral visual examination on all the members of the households above the age of 10 years" Characteristics and proportion of individuals who did not receive the index test(s) and/or reference standard or excluded from analysis: From 48,080 participants initially eligible, 5761 unavailable for examination by reference standard, and a further 7553 "who did not comply with the study procedure were excluded from the study population." Results available for 34,766 participants (38% attrition) Characteristics and proportion of individuals who received a reference standard other than examination and clinical evaluation by a specialist physician: Reference standard carried out by a trained health worker
Comparative	
Notes	Possible bias introduced through exclusion of participants that did not comply with the procedure. Participants located in area of high prevalence of oral cancer and potentially malignant lesions
Methodological quality	

Elango 2011 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it prespecified?	Yes		
Was conflict of interest avoided?	Yes		
Where multiple index tests were used, were the results of the second index test interpreted without knowledge of the results of the first index test?			
		Unclear	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Unclear**
DOMAIN 4: Flow and Timing			

Elango 2011 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
	High

Ikeda 1995

Study characteristics	
Patient sampling	Method of patient selection: Postal invitation to 60-year old residents to participate in an annual mass screening programme
Patient characteristics and setting	154 from last screening exercise (5187 eligible during reported 7 years of the programme from 1986 to 1993) Age: 60 years of age Sex: Not stated SES: Not stated Ethnicity: Not stated Stated risk factors: Not stated Previous history: Not stated Location: Japan Clinical setting: City health centre
Index tests	Index test: Standard visual examination carried out by 4 general dental practitioners. "Lesions were recorded on a standard WHO form modified for local conditions" Description of positive case definition by index test as reported: "The screen was recorded as positive for oral cancer or precancer if the examiner considered a carcinoma, erythroplakia, lichen planus or chronic candidosis was present." Types of lesion categorised as malignancy, malignant potential, benign characterisation or absence Sequence of tests: Index followed by reference. "Following screening individual consultation was provided on site for all those examined..." Training or calibration: Trained according to WHO guidelines. Calibration for the 4 dentists was reported. Kappa scores were slight to moderate (0.08 to 0.44) for classification of lesions and moderate to substantial (0.39 to 0.78) for identifying the presence/absence of lesions Blinding of examiners: Index test completed prior to reference Conflict of interests: Not stated

Ikeda 1995 *(Continued)*

Target condition and reference standard(s)	Target condition: As for index test. Reference standard: "..assessed by an oral medicine specialist" Description of positive case definition by reference test as reported: Presence or absence of malignant or pre-malignant oral lesions and classification of lesions Training or calibration: Previously calibrated, details not reported Blinding of examiners: Not explicitly stated but "...independent clinical diagnoses of the instructor carried out concurrently" Prevalence of the target condition in the sample: 15/154 9.7%
Flow and timing	Time interval and any interventions between index test(s) and reference standard: "Following screening..." consultation undertaken on same day Characteristics and proportion of individuals who did not receive the index test(s) and/or reference standard or excluded from analysis: All received index and reference (data fully reported for results of most recent screening exercise only) Characteristics and proportion of individuals who received a reference standard other than examination and clinical evaluation by a specialist physician: Screened positive did receive biopsy but data taken from table of clinical diagnosis of specialist (Table 1)
Comparative	
Notes	Definition of positive threshold could underestimate accuracy 802/5187 eligible residents presented for screening during reported 7 years of the programme from 1986 to 1993

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was conflict of interest avoided?	Yes		

Ikeda 1995 *(Continued)*

Where multiple index tests were used, were the results of the second index test interpreted without knowledge of the results of the first index test?			
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Julien 1995

Study characteristics	
Patient sampling	Method of patient selection: Participants recruited "...by the screener or the specialist from the various outpatient departments of the hospital." Method of selection of participants at the dental hospital is unclear
Patient characteristics and setting	1042 participants (total population not reported) Participant characteristics are reported across both studies. Age: 40 years or over; mean 56 years Sex: 892 male 1135 female SES: Not stated Ethnicity: Not stated

Julien 1995 *(Continued)*

	Stated risk factors: 162 heavy smoker 608 moderate smoker 1257 non-smoker; 61 heavy drinker 527 moderate drinker 1439 light drinker Previous history: Not stated Location: UK Clinical setting: Out-patient departments of a dental hospital
Index tests	Index test: Thorough visual examination of the surface of the oral mucosa according to the British Postraduate Medical Federation, 1991, by either a general dental practitioner, a community dental officer or a junior hospital dentist (24 screeners) Description of positive case definition by index test as reported: "A lesion was defined as positive when a white patch, red patch, or an ulcer of longer than two weeks duration was detected." "The screeners were also instructed to include lesions of lupus erythematosus, submucous fibrosis or actinic keratosis as positive." All types of lichen planus were also regarded as positive Sequence of tests: Index followed by reference Training or calibration: "..screeners advised of diagnostic criteria which should result in a positive or negative screen .....no formal training or standardisation was undertaken" Blinding of examiners: Index test completed before reference Conflict of interests: Supported by grant from the Department of Health, UK
Target condition and reference standard(s)	Target condition: Oral cancer and pre-cancer Reference standard: Visual examination by second dental specialist who was able to refer subjects for further tests or review as appropriate (single specialist) Description of positive case definition by reference test as reported: As for index test. "A lesion was defined as positive when a white patch, red patch, or an ulcer of longer than two weeks duration was detected." "The screeners were also instructed to include lesions of lupus erythematosus, submucous fibrosis or actinic keratosis as positive." All types of lichen planus were also regarded as positive Training or calibration: Not stated but quoted as "a specialist." Single examiner so no calibration Blinding of examiners: Index test completed before reference. "The results were also recorded on a standard form which was collated with the screeners' form only after completion." "All subjects were examined by a specialist who provided an independent definitive diagnosis" Prevalence of the target condition on the sample: 32/1042 3.1%
Flow and timing	Time interval and any interventions between index test(s) and reference standard: Not explicit, however, reasonable to assume both conducted on same visit Characteristics and proportion of individuals who did not receive the index test(s) and/or reference standard or excluded from analysis: None Characteristics and proportion of individuals who received a reference standard other than examination and clinical evaluation by a specialist physician: None
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

Julien 1995 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
		Unclear	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was conflict of interest avoided?	Yes		
Where multiple index tests were used, were the results of the second index test interpreted without knowledge of the results of the first index test?			
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		

Julien 1995 *(Continued)*

Were all patients included in the analysis?	Yes

Low

Julien 1995a

Study characteristics	
Patient sampling	Method of patient selection: List of registered medical practice patients obtained and postal invitation to participate in screening sent
Patient characteristics and setting	985 participants (total population not reported) Participant characteristics are reported across both studies. Age: 40 years or over Sex: 892 male 1135 female SES: Not stated Ethnicity: Not stated Stated risk factors: 162 heavy smoker 608 moderate smoker 1257 non-smoker; 61 heavy drinker 527 moderate drinker 1439 light drinker Previous history: Not stated Location: UK Clinical setting: Inner city medical practice
Index tests	Index test: Thorough visual examination of the surface of the oral mucosa according to the British Postraduate Medical Federation, 1991, by either a general dental practitioner, a community dental officer or a junior hospital dentist (24 screeners) Description of positive case definition by index test as reported: "A lesion was defined as positive when a white patch, red patch, or an ulcer of longer than two weeks duration was detected." "The screeners were also instructed to include lesions of lupus erythematosus, submucous fibrosis or actinic keratosis as positive." All types of lichen planus were also regarded as positive Sequence of tests: Index followed by reference Training or calibration: "..screeners advised of diagnostic criteria which should result in a positive or negative screen .....no formal training or standardisation was undertaken" Blinding of examiners: Index test completed before reference Conflict of interests: Supported by grant from the Department of Health, UK
Target condition and reference standard(s)	Target condition: Oral cancer and pre-cancer Reference standard: Visual examination by second dental specialist who was able to refer subjects for further tests or review as appropriate (single specialist) Description of positive case definition by reference test as reported: As for index test. "A lesion was defined as positive when a white patch, red patch, or an ulcer of longer than two weeks duration was detected." "The screeners were also instructed to include lesions of lupus erythematosus, submucous fibrosis or actinic keratosis as positive." All types of lichen planus were also regarded as positive

Julien 1995a *(Continued)*

	Training or calibration: Not stated but quoted as "a specialist". Single examiner so no calibration Blinding of examiners: Index test completed before reference. "The results were also recorded on a standard form which was collated with the screeners' form only after completion." "All subjects were examined by a specialist who provided an independent definitive diagnosis" Prevalence of the target condition on the sample: 22/985 2.2%
Flow and timing	Time interval and any interventions between index test(s) and reference standard: Not explicit, however, reasonable to assume both conducted on same visit Characteristics and proportion of individuals who did not receive the index test(s) and/or reference standard or excluded from analysis: None Characteristics and proportion of individuals who received a reference standard other than examination and clinical evaluation by a specialist physician: None
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was conflict of interest avoided?	Yes		
Where multiple index tests were used, were the results of the second index test interpreted without knowledge of the results of the first index test?			
		Low	**Low**
DOMAIN 3: Reference Standard			

Julien 1995a *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Mathew 1997

Study characteristics	
Patient sampling	Method of patient selection: Re-examination of 2069 eligible participants from the 9000 participants recruited in January to May 1996, shortly after commencement of the study. "Subjects were selected by choosing densely inhabited areas to allow re-examination of as many subjects as possible in two weeks." Study looking at the reproducibility and validity of oral visual inspection by health workers within a randomised controlled intervention trial of visual screening
Patient characteristics and setting	2069 participants Age: Mean 47.7 years, sd 9.1 years (range 35 to 64 years) Sex: 678 males; 1391 females SES: Recorded but not reported Ethnicity: Recorded but not reported Stated risk factors: Details on smoking and alcohol were recorded but not reported Previous history: Recorded but not reported Location: Kerala, India Clinical setting: Participants' homes
Index tests	Index test: Systematic oral visual examination by trained health workers (n = 14) in the inspection and detection of oral lesions

Mathew 1997 *(Continued)*

	Description of positive case definition by index test as reported: "...homogeneous leucoplakia, ulcerated leucoplakia, verrucous leucoplakia, erythroplakia, nodular leukoplakia, submucous fibrosis, and oral cancer" Sequence of tests: Initial screen by health worker followed by second screen (the index test) by same health worker (1 to 6 months later) to establish reliability. 2069 received the index test (second screen by HW) and this formed the sample for the sensitivity and specificity calculations Training or calibration: "Training sessions spread over 6 weeks composed of lectures, practical demonstrations and field work conducted by Faculty... At the end of training sessions written and practical tests were conducted identifying the best health workers.... They were also given manuals and photographic documentation to identify different types of oral lesions." The "best performing" health workers were retained for the study Blinding of examiners: Index test completed before reference Conflict of interests: Supported by a grant from the Association of International Cancer Research, Scotland, UK
Target condition and reference standard(s)	Target condition: As for index test "...homogoneous leukoplakia, ulcerated leukoplakia, verrucous leukoplakia, erythroplakia, nodular leukoplakia, submucous fibrosis, and oral cancer" Reference standard: Visual examination by a specialist physician (decision made by single physician, 1 of 3). "....comparison with pathological findings is not possible as biopsy has not been performed for most case. Biopsy is performed for cases of nodular leucoplakias, erythroplakias and suspicious growths only, and this is currently being undertaken" Training or calibration: 100 participants formed the basis of comparability of findings evaluation. Kappa value of 0.85 was reported for the findings of the 3 physicians Blinding of examiners: Reference test undertaken immediately after index test. Both health worker and specialist in participants' home at the same visit Prevalence of the target condition on the sample: 212/2069 10.3%
Flow and timing	Time interval and any interventions between index test(s) and reference standard: "This was immediately followed by an independent examination of the same subject by one of three physicians" Characteristics and proportion of individuals who did not receive the index test(s) and/or reference standard or excluded from analysis: None Characteristics and proportion of individuals who received a reference standard other than examination and clinical evaluation by a specialist physician: None
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		

Mathew 1997 *(Continued)*

		Low	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was conflict of interest avoided?	Yes		
Where multiple index tests were used, were the results of the second index test interpreted without knowledge of the results of the first index test?			
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Mehta 1986

Study characteristics	
Patient sampling	Method of patient selection: For the screening study, a basic health worker visited each household to report on health status in an area of high oral cancer prevalence. "Four adjacent blocks, two as study area I (pop 218728) and two as study area II (pop 250,399) were selected for this investigation." Field checking of the diagnosis of the health worker by the study dentist was initiated after 6 months and completed for 40 health workers. For each of the health workers' lists "A house with a lesion case was selected as a nodal point and all the available individuals from nearby houses who figured in the list were examined." Carried out on high risk individuals within a household "..i.e. people aged 35 years and above with tobacco habits"
Patient characteristics and setting	2063 'high risk' participants (out of 39,331 screened) Age: 35 years and above Sex: Not stated SES: Not stated Ethnicity: Not stated Stated risk factors: All participants had tobacco habits, HPV and alcohol use not reported Previous history: Not stated Location: Kerala, India Clinical setting: Participants' homes
Index tests	Index test: Standard visual examination by basic health worker working to a reference manual Description of positive case definition by index test as reported: Referable lesions were "nodular leukoplakia, submucous fibrosis, and ulcers and growths suggestive of oral cancer." Non-referable lesions included "homogenous leukoplakia, oral lichen planus, smoker's palate and central papillary atrophy of the tongue papillae." Definition of positive threshold may over-estimate accuracy values (homogenous leukoplakia considered to be test negative) Sequence of tests: Index followed by reference Training or calibration: Yes. Training provided by dentists, members of the research team. "The final performance of the trainees was judged as satisfactory" Blinding of examiners: Index test completed before reference Conflict of interests: None stated. Study was supported by a grant from the National Institutes of Health
Target condition and reference standard(s)	Target condition: Referable lesion Reference standard: Standard visual examination by dentist (member of research team) in participants' home Description of positive case definition by reference test as reported: Referable lesions were "nodular leukoplakia, submucous fibrosis, and ulcers and growths suggestive of oral cancer." Non-referable lesions included "homogenous leukoplakia, oral lichen planus, smoker's palate and central papillary atrophy of the tongue papillae." Definition of positive threshold may over-estimate accuracy values (homogenous leukoplakia considered to be test negative) Training or calibration: The research team of dentists "...was experienced in conducting house to house surveys for oral cancer and precancerous lesions in rural areas of Ernakulam district for 16 years" Blinding of examiners: Unclear whether the dentists were aware of the screening results. "The list contained the categorization indicated by the BHW"

Mehta 1986 *(Continued)*

	Prevalence of the target condition on the sample: 27/1921 1.41%
Flow and timing	Time interval and any interventions between index test(s) and reference standard: At the same visit. "One day was devoted to rechecking for each of the 40 BHW" Characteristics and proportion of individuals who did not receive the index test(s) and/or reference standard or excluded from analysis: 142 were falsely reported to have been examined by the BHW, and they were excluded from further analysis. Exclusions are unlikely to induce bias Characteristics and proportion of individuals who received a reference standard other than examination and clinical evaluation by a specialist physician: None
Comparative	
Notes	Data presented for field check only, not full screening programme

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was conflict of interest avoided?	Yes		
Where multiple index tests were used, were the results of the second index test interpreted without knowledge of the results of the first index test?			
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly	Yes		

Mehta 1986 *(Continued)*

classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
		Low	

Scott 2010

Study characteristics	
Patient sampling	Method of patient selection: "Participants were recruited from a general practitioner's list in South East London, UK. Patients who were at risk of oral cancer (aged 45 years or older and who smoked) were identified as potential participants by their general practitioner." Recruitment was by invitation letter to 243 eligible patients. 53 patients participated
Patient characteristics and setting	53/243 eligible patients Age: Mean age 54 years (sd 5.9 years, range 45 to 64 years) Sex: 36 male 17 female SES: 24 no/compulsory education; 25 beyond compulsory education Ethnicity: 37 white 14 other Stated risk factors: 40 hazardous drinking (AUDIT-C) 11 alcohol dependent; 41 current smoker 12 used to smoke; 27 regular attenders, 10 irregular attenders, 15 emergency or never Previous history: Not stated Location: South East London, UK Clinical setting: "Research room"
Index tests	Index test: Mouth self examination in accordance with a patient leaflet, at the same location. "The leaflet had been specifically developed for and piloted with heavy smokers and drinkers and has a reading age of 10 to 12 years and a Flesch reading ease score of 79%, indicating it can be read and understood with ease"

Scott 2010 *(Continued)*

	Description of positive case definition by index test as reported: Red patches, white patches, ulcers and lumps or swellings Sequence of tests: Reference followed by index Training or calibration: Conducted mouth self examination in accordance with specifically developed patient leaflet Blinding of examiners: Reference preceded index test. "After the dentist's examination (yet before the results of the examination were revealed to the participant)..." "The dentist remained in the room but did not assist the participant in conducting the mouth self examination" Conflict of interests: The study was funded by a Cancer Research UK Pilot Project Award (C19770/A8554), but no conflict of interest
Target condition and reference standard(s)	Target condition: Red patches, white patches, ulcers and lumps or swelling Reference standard: Examination by single dentist (member of research team). Protocol for examination reported Description of positive case definition by reference test as reported: "The presence (including site and provisional diagnosis) and absence of potentially malignant oral lesions (ulcers, white or red patches, or lumps/swellings) were noted on a pro forma" Training or calibration: Experience and training not reported Blinding of examiners: Yes. Reference standard proceeded index test Prevalence of the target condition on the sample: 12/53 22.6%
Flow and timing	Time interval and any interventions between index test(s) and reference standard: Reference test immediately followed index test Characteristics and proportion of individuals who did not receive the index test(s) and/or reference standard or excluded from analysis: None Characteristics and proportion of individuals who received a reference standard other than examination and clinical evaluation by a specialist physician: None
Comparative	
Notes	Low response rate for participation 53/243 eligible patients recruited from an "at risk" group

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowl-	Unclear		

Scott 2010 *(Continued)*

edge of the results of the reference standard?			
If a threshold was used, was it pre-specified?	Yes		
Was conflict of interest avoided?	Yes		
Where multiple index tests were used, were the results of the second index test interpreted without knowledge of the results of the first index test?			
		Unclear	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Su 2010

Study characteristics	
Patient sampling	Method of patient selection: Community-based randomised controlled trial of Toluidine blue for the detection and incidence of oral cancer. Mass screening programme (eligible at 15 years old or over) aimed at detecting 5 prevalent neoplasms (cervical, breast, hepatocellular, colorectal, and oral cancer) and 3 chronic diseases (hypertension, diabetes, and hyperlipidaemia). From the mass screening programme individuals were ineligible for the randomised controlled trial if they "lacked oral habits such as cigarette smoking or chewing betel quid." Randomised to either visual examination plus Toluidine blue (experimental group) or to visual examination alone (control group)

Su 2010 *(Continued)*

Patient characteristics and setting	Analysis of data of 7975 participants enrolled into the randomised controlled trial during 2000 Age: Mean 44.9 years sd 14.4; mean 44.6 years sd 15.3 Sex: Male 3719 and 3550; female 361 and 345 SES: Not stated Ethnicity: Not specified Stated risk factors: Participants were smokers or betel quid chewers, HPV or alcohol consumption not reported Previous history: Not stated Location: Taiwan Clinical setting: Randomised controlled trial as part of community-based screening programme
Index tests	Target condition: Asymptomatic oral pre-malignant lesions (OPML) and oral cancer. Oral submucous fibrosis, homogenous leukoplakia, non-homogeneous leukoplakia, erythroplakia and oral cancer Index test (2): Visual examination by dentist plus Toluidine blue (experimental group) Visual examination by dentist alone (control group) Description of positive case definition by index test as reported: "The presence of any visible lesion in the oral cavity was recorded as screen-positive." Information reported for screen positive rate and detection rate Sequence of tests: Index test followed by reference standard Training or calibration: Training given to dentists was carried out by a senior oral pathologist. No calibration was reported Blinding of examiners: Index test followed by reference standard. Placebo dye Conflict of interests: None declared
Target condition and reference standard(s)	Target condition: Any visible lesion (detection), oral cancer (incidence rate of oral cancer, diagnostic accuracy) Reference standard: Only screened positives referred for biopsy; entire cohort (screened positive or screened negative) assessed through national cancer registry Description of positive case definition by reference test as reported: As indicated by national cancer registry Training or calibration: "Diagnostic criteria, examination procedures, and documentation formats were discussed, taught, and calibrated in advance for all personnel participating in the study" Blinding of examiners: All personnel were unaware of group allocation Prevalence of the target condition on the sample: 0.12% and 0.15% in each trial arm
Flow and timing	Time interval and any interventions between index test(s) and reference standard: Screened positive participants were referred for a definite clinical diagnosis within 10 to 14 days. 5-year follow-up of oral cancer development through linkage to the national cancer registry Characteristics and proportion of individuals who did not receive the index test(s) and/or reference standard or excluded from analysis: None

Su 2010 *(Continued)*

	Characteristics and proportion of individuals who received a reference standard other than examination and clinical evaluation by a specialist physician: All. Quote: "We retrieved the occurrence of oral cancer, survival status, and causes of death of the studied participants by linking the entire cohort with the National Cancer Registry and the National Household Registry until December 31, 2004"
Comparative	
Notes	Estimates of sensitivity and specificity of the index tests are based on the outcome of oral cancer as indicated by the national cancer registry. Results presented for detection rate ratio for oral pre-malignant lesions and malignant lesions and incidence rate of oral cancer

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was conflict of interest avoided?	Yes		
Where multiple index tests were used, were the results of the second index test interpreted without knowledge of the results of the first index test?			
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Su 2010 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear

		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
		Low	

Sweeny 2011

Study characteristics	
Patient sampling	Method of patient selection: "...a prospective study was performed at the University of Alabama at Birmingham. Consecutive patients who presented to the Otolaryngology clinic between November 2009 and October 2010 for follow-up (n = 88) following management of primary head and neck cancer"
Patient characteristics and setting	88 participants Age: Mean 64 years (range 41 to 85 years) Sex: 65 male 23 female SES: Not reported Ethnicity: 54 Caucasian Stated risk factors: 58 alcohol consumption; 71 history of tobacco use Previous history: "All patients had undergone a previous treatment for head and neck cancer." "All patients evaluated during routine surveillance visits" Location: Alabama, USA Clinical setting: Otolaryngology clinic
Index tests	Index test (3): "..sites were initially screened by a registered nurse and then by a fellowship trained head and neck surgeon using visualization with white light illumination (traditional exam light) followed by visualization of tissue autofluorescence and tissue reflectance. The Trimira® Identafi® 3000 ultra, multi-spectral oral cavity screening system was used." "Patients were evaluated by direct visualization of the oral cavity with white light (traditional exam light), tissue autofluorescence and tissue reflectance." Only the results of visualisation examination with white light are included in this analysis as the autofluorescence and reflectance data are not presented as adjuncts but as independent tests

Sweeny 2011 *(Continued)*

	Description of positive case definition by index test as reported: "oral cavity cancer." Abnormality/lesion with concern for malignancy or recurrence. Not explicitly stated Sequence of tests: Index followed by reference Training or calibration: Not stated but index test conducted by registered nurse followed by head and neck surgeon Blinding of examiners: Not stated but index tests preceded reference test. No information of blinding after successive index tests Conflict of interests: This work was supported by a grant from the National Institute of Health (2T32 CA091078-09), but no conflict of interest
Target condition and reference standard(s)	Target condition: Head and neck cancer recurrence Reference standard: "Screening results were compared to histological biopsy results or a three month follow-up screening. Any area of abnormality found by visualization with traditional white light illumination and/or by tissue autofluorescence or reflectance was biopsied and evaluated by a pathologist using standard histopathologic analysis" Description of positive case definition by reference test as reported: "Positive disease" Training or calibration: Not stated Blinding of examiners: Not stated Prevalence of the target condition on the sample: 4/88 4.6%
Flow and timing	Time interval and any interventions between index test(s) and reference standard: Not explicitly stated. Follow-up screening visit at 3 months Characteristics and proportion of individuals who did not receive the index test(s) and/or reference standard or excluded from analysis: None Characteristics and proportion of individuals who received a reference standard other than examination and clinical evaluation by a specialist physician: Biopsy for screened positive participants. Reference standard by follow-up visit for some participants (number of participants not specified)
Comparative	
Notes	"Our study was unique in that it evaluated the population most likely to benefit from screening." Participants attending for routine surveillance

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**High**
DOMAIN 2: Index Test All tests			

Sweeny 2011 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was conflict of interest avoided?	Yes		
Where multiple index tests were used, were the results of the second index test interpreted without knowledge of the results of the first index test?	Unclear		
		Unclear	**Unclear**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Unclear**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
		Unclear	

Warnakulasuriya 1990

Study characteristics	
Patient sampling	Method of patient selection: Screening programme at a rural location, Kadugannawa, Sri Lanka. "The PHC workers carried out an examination….of people over the age of 20 years in their area;….voters lists were used to identify and record the persons examined and those who were referred"
Patient characteristics and setting	Population of 87,277 adults (> 20 years of age) of whom 29,295 were screened during study periods of 52 weeks. From this number 1872 received both the index test and the reference test. Patient characteristic information reported only for those screened positive and attending the referral centre Age: 20 to 39 years n = 182, 40 to 59 years n = 315, > 60 n = 163 Sex: 480 male 180 female SES: Not stated Ethnicity: Not stated Stated risk factors: Not stated Previous history: Not stated Location: Sri Lanka Clinical setting: Participants' own homes
Index tests	Index test: Examination of the lining mucosa of the oral cavity in natural daylight using dental mirrors by primary health care (PHC) workers comprising midwives, public health inspectors and public health nurses Description of positive case definition by index test as reported: "The PHC workers identified positive cases on the basis of simple, explicitly stated criteria. The diagnosis criteria included the presence of a white or red lesion on the oral mucosa with a smooth, corrugated or nodular surface which cannot be scraped of using the dental mirror head. Elevated and ulcerated areas with co-existing red or white lesions were also referable" Sequence of tests: Index followed by reference Training or calibration: "...participated in a two-day training programme which provided a clinical demonstration of oral cancer and precancer, instructions regarding the screening methods and referral mechanisms" Blinding of examiners: Index test preceded reference test Conflict of interests: Authors declare no conflict of interest. Work was supported by the Cancer Control Programme of Sri Lanka
Target condition and reference standard(s)	Target condition: Oral cancer/pre-cancer (for purposes of accuracy of examination). Leukoplakia, erythroplakia or carcinoma Reference standard: Re-examination by the project dentist Description of positive case definition by reference test as reported: Oral cavity cancer Training or calibration: Not stated but carried out by experienced dentists Blinding of examiners: Unclear. Re-examination of screened positive cases took place at the referral centre "(all screened positives were referred); a sample of screened negative participants were randomly selected from PHC files by the project dentist visiting each field area" Prevalence of the target condition on the sample: 405/1872 21.6% (sample for diagnostic test accuracy assessment), 660/29,295 screened positive referable lesions 2.25%
Flow and timing	Time interval and any interventions between index test(s) and reference standard: Re-examination of "660 cases who arrived at the referral centre within 18 months (January 1981 to June 1982) after case detection."

Warnakulasuriya 1990 *(Continued)*

	"...negative cases randomly selected from PHC files.. were re-examined, during the three month period of initial PHC examinations" Characteristics and proportion of individuals who did not receive the index test(s) and/or reference standard or excluded from analysis: 87,277 adults were eligible for the screening programme of whom 29,295 were screened. "All referred (screened positive) participants who arrived at the referral centre were re-examined by the project dentist to validate the PHC diagnosis." "A sample of negative cases was randomly selected from PHC files (in whom PHC workers had not recorded a lesion) were re-examined, during the three month period of initial examination. A minimum of 30 negative cases from each PHC file were thus re-examined." 1872 received both the index test and the reference test Characteristics and proportion of individuals who received a reference standard other than examination and clinical evaluation by a specialist physician: None
Comparative	
Notes	Only 660 screened positive participants arrived at the referral centre within 18 months after screen positive detection; 54.1% of detected cases in the field Index test target condition "white or red lesion that cannot be scraped off"; reference standard for accuracy of screening "correctly referred cases who, on examination, had oral cancer or precancer" Prevalence in sample for diagnostic test accuracy assessment was high 21.6%

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it prespecified?	Yes		
Was conflict of interest avoided?	Yes		
Where multiple index tests were used, were the results of the second in-			

Warnakulasuriya 1990 *(Continued)*

dex test interpreted without knowledge of the results of the first index test?			
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
		High	

Warnakulasuriya 1991

Study characteristics	
Patient sampling	Method of patient selection: Optional screening programme at a rural location, Galle, Sri Lanka. Primary health care (PHC) workers carried out a visual oral examination of people over the age of 20 years in their geographical area. The 1981 electoral list was used to identify eligible individuals
Patient characteristics and setting	Population of 72,867 adults (> 20 years of age) of whom 57,124 were examined during 1 year by PHC workers. From this number 3543 received both the index test and the reference test Age: Participants were 20 years of age or older Sex: Not stated

Warnakulasuriya 1991 *(Continued)*

	SES: Not stated Ethnicity: Not stated Stated risk factors: Not stated Previous history: Not stated Location: Sri Lanka Clinical setting: Participants' own homes
Index tests	Index test: Examination of the lining mucosa of the oral cavity in natural daylight using dental mirrors by PHC workers Description of positive case definition by index test as reported: "The PHC workers identified positive cases on the basis of simple, explicitly stated criteria. The diagnosis criteria included the presence of a white or red lesion on the oral mucosa with a smooth, corrugated or nodular surface which cannot be scraped of using the dental mirror head. Elevated and ulcerated areas with co-existing red or white lesions were also referable" Sequence of tests: Index test followed by reference test Training or calibration: Participated in a 2-day training programme which provided a clinical demonstration of oral cancer and pre-cancer, instructions regarding the screening methods and referral mechanisms, as in the pilot study (Warnakulasuriya 1990) Blinding of examiners: Index test followed by reference test Conflict of interests: Authors declare no conflict of interest. Work was supported by funds from the National Cancer Control Programme of Sri Lanka
Target condition and reference standard(s)	Target condition: Oral cancer/pre-cancer (for purposes of accuracy of examination) Reference standard: Re-examination by the project dentist. "The hospital dental surgeon reexamined all referred subjects to revalidate the diagnosis given by the PHCW." "Biopsies were obtained from all cases suggestive of oral cancer and a representative sample of precancers was also made by incision biopsy" Description of positive case definition by reference test as reported: Oral cavity cancer Training or calibration: "A hospital dentist attached to a local hospital and who had received special training in oral cavity examinations was assigned to supervise the project" Blinding of examiners: Unclear. Re-examination of screened positive cases took place at the referral centre (all screened positives were referred); a sample of screened negative participants were randomly selected from PHC files by the project dentist visiting each field area Prevalence of the target condition on the sample: 1797/3543 50.7% (sample for diagnostic test accuracy assessment); 3559/57,124 6.23% screened positive (oral lesions)
Flow and timing	Time interval and any interventions between index test(s) and reference standard: Not stated Characteristics and proportion of individuals who did not receive the index test(s) and/or reference standard or excluded from analysis: 72,867 adults were eligible for the screening programme of whom 57,124 were screened. Re-examination of 2193 participants who arrived at the referral centre out of 3559 who screened positive. Field checking of 1350 screened negative cases was undertaken (random sample from electoral list). 21 excluded from analysis due to non-diagnosis from PCH worker. 3543 participants received both the index test and the reference test Characteristics and proportion of individuals who received a reference standard other than examination and clinical evaluation by a specialist physician: None
Comparative	

Warnakulasuriya 1991 *(Continued)*

Notes	Only 2193 screened positive participants arrived at the referral centre; 62% of detected cases in the field Prevalence in sample for diagnostic test accuracy assessment was very high 50.7%

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was conflict of interest avoided?	Yes		
Where multiple index tests were used, were the results of the second index test interpreted without knowledge of the results of the first index test?			
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		

Warnakulasuriya 1991 *(Continued)*

		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
		Unclear	

HPV = human papillomavirus; sd = standard deviation; SES = socio-economic status

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Bhalang 2008	Patients suspected of oral squamous cell carcinoma
Bowles 1973	Patients suspected of cancer
Chen 2007	Presenting with lesions
Csépe 2007	Prevalence data and risk factors
Fernández Garrote 1995	Data on referral, incidence and stage
Hapner 2011	Prevalence data
Huber 2004	Exploration of oral soft tissue under chemiluminescent illumination
Huff 2009	Inappropriate study design
Leocata 2007	Prevalence data
Lim 2003	Prevalence data
Nagao 2000	Participation rates and prevalence data; no screen negatives verified
Oh 2007	Outcomes measured on a lesion level. Cross-tabulation table cannot be constructed
Poh 2007	Prevalence data
Srivasta 1971	Chronic ulcerative lesions
Vahidy 1972	Presenting with lesions
Warnakulasuriya 2010	Prevalence data

Characteristics of ongoing studies *[ordered by study ID]*

Kulak 2010

Trial name or title	The use of autofluorescence in detection of oral lesions
Target condition and reference standard(s)	Oral lesions Head and neck examination under standard light
Index and comparator tests	Fiberoptic examination VELscope exam
Starting date	Not stated. Publication in 2010 with results for 17/300 participants required
Contact information	Jessica Kulak, jkulak2@med.miami.edu
Notes	

DATA

Presented below are all the data for all of the tests entered into the review.

Table Tests. Data tables by test

Test	No. of studies	No. of participants
1 Conventional oral examination	10	25568
2 Mouth self examination	2	34819

Test 1. Conventional oral examination.

Test 2. Mouth self examination.

ADDITIONAL TABLES

Table 1. Screening tests for PMDs and oral cavity cancer

Test	Characteristics	Classification of response	Other information
Conventional oral examination (COE)	A standard visual and tactile examination of the oral mucosa under normal (incandescent) light	The presence of an oral mucosal abnormality is classified as a positive	Traditionally been used as an oral cancer screen, but its utility is debated (Lingen 2008)

Table 1. Screening tests for PMDs and oral cavity cancer *(Continued)*

		test result; the absence of any oral mucosal abnormalities is classified as a negative test result	**Advantages**: quick and easy once trained, minimally invasive **Disadvantages**: oral mucosal abnormalities are not necessarily clinically or biologically malignant; only a small percentage of leukoplakias are progressive or become malignant; COE cannot distinguish between those that are or are not; some precancerous lesions may exist within oral mucosa that appears clinically normal by COE alone (Lingen 2008)
Vital rinsing (e.g. Toluidine blue, Tolonium chloride)	Vital rinsing refers to the use of dyes such as Toluidine blue or Tolonium chloride to stain oral mucosa tissues for PMD or malignancy (Leston 2010; Lingen 2008; Patton 2008). The procedure is as follows • Pre-rinse with acetic acid • Rinse with water • Apply Toluidine blue • Post-rinse with acetic acid • Rinse with water • Observe mucosa to check for staining	The result of the test is classified as positive if tissue is stained and negative if no tissue is stained, or equivocal if no definitive result can be obtained	**Advantages**: ability to define areas that could be malignant or abnormal but cannot be seen; assess the extent of the PMD for excision **Disadvantages**: benign inflammatory lesions subject to stain; failure of some cancerous lesions to stain; variation in test performance depending on how thorough the test procedures are followed; contraindicated in those who are known to be allergic to iodine
Light-based detection (e.g. ViziLite and ViziLite plus, Microlux/DL, VELscope, Identafi 3000)	Light-based systems to identify pre-malignant and malignant lesions and to highlight their presence through tissue autofluorescence or reflectance (Leston 2010; Lingen 2008; Patton 2008). E.g. using ViziLite Plus or Microlux/DL, the procedure is as follows (Lingen 2008) • Pre-rinse with acetic acid • Use blue-light source to visually assess the oral cavity ViziLite Plus also provides a tolonium chloride solution (TBlue) to aid in the marking of the lesion for biopsy once the light source is removed	The result of the test is classed as negative if the appearance of the epithelium is lightly bluish white and positive if the appearance of the epithelium is distinctly white (acetowhite) For systems based on autofluorescence the result of the test is classed as negative if fluorescence is maintained and positive if fluorescence is lost	**Advantages**: simple to use; non-invasive; do not require consumable re-agents; provide real time results; can be performed by a wide range of operators after a short training period **Disadvantages**: the necessity of a dark environment; high initial set up (for VELscope) or recurrent costs (for ViziLite in low-income countries); lack of permanent record unless photographed; inability to objectively measure visualisation results
Mouth self examination	Self examination, usually in the home setting in accordance with instructional material	Usually the presence of any lesion	**Advantages**: simple to carry out and low cost. Can be carried out in an individual's own home **Disadvantages**: target condition is the presence or absence of oral lesions. Cannot differentiate between potentially malignant and non-malignant lesions
Blood and saliva analyses	These novel technologies are at an early stage of development and evaluation Analysis of blood or saliva samples which tests for the presence	Cut-off probabilities vary widely and are dependent on the individual bio-marker or combina-	**Advantages**: non-invasive (saliva tests) or minimally invasive (blood tests) **Disadvantages**: there is a tendency for the estimated diagnostic accuracy of new health technologies to decline over time as evidence from indepen-

Table 1. Screening tests for PMDs and oral cavity cancer *(Continued)*

	of bio-markers of PMD and oral cancer (Brinkmann 2011; Lee 2009; Li 2006)	tion of bio-markers examined Molecular markers for diagnosis include changes in cellular DNA, altered mRNA transcripts, altered protein levels	dent evaluations accumulate (Wyatt 1995). This bias, which can be substantial, has been demonstrated in other domains, e.g. acute abdominal pain (Liu 2006) and clinical decision support systems (Garg 2005). Promising bio-marker tests in several clinical areas were eventually been shown to be disappointing (Buchen 2011). It remains to be seen whether this is the case with oral cancer and PMDs

PMDs = potentially malignant disorders

Table 2. Indicators for the assessment of methodological quality

Domain	Patient selection	Index test	Reference standard	Flow and timing
Description	Describe methods of patient selection. Describe included patients (characteristics, prior testing, presentation, intended use of index test and setting)	Describe the index test and how it was conducted and interpreted. Describe the sequence of tests, any training or calibration of assessors (levels of agreement should be reported. Where this is measured by the kappa statistic*, acceptable values range from 0.61 (moderate agreement) to 1.00 (almost perfect agreement) (Landis 1977)), any procedures taken to ensure blinding of examiners, post-hoc or a priori threshold specification, any conflict of interest or commercial funding *This statistic is a measure of inter-rater agreement of observations measured at a categorical level	Describe the reference standard and how it was conducted and interpreted. Any measures taken to ensure assessors were blinded to the results of the index tests should be documented, along with the sequence of reference and index tests	Describe the characteristics and proportion of patients who did not receive the index test(s) and/or reference standard, who received a reference standard other than examination and clinical evaluation by a specialist physician, or who were excluded from the 2 x 2 table (refer to flow diagram). Describe the time interval and any interventions between index test(s) and reference standard. The length of time between the index test and reference standard should be short in the majority of cases. If the period elapsed between initial screening and reference standard (examination and clinical evaluation) is greater than 6 weeks then this was considered an unacceptable delay
Signalling questions (Yes/No/Unclear)	Was a consecutive or random sample of patients enrolled? Classify as **Yes** if consecutive patients or a random sample of individuals were recruited Classify as **No** if non-consecutive patients or a non-ran-	Were the index test results interpreted without knowledge of the results of the reference standard? Classify as **Yes** if interpreters of index test results clearly do not know results of reference standard Classify as **No** if interpreters of index test results clearly know results of reference standard Classify as **Unclear** if study did not provide any information on whether interpreters of index tests were blinded to reference standard	Is the reference standard likely to correctly classify the target condition? The reference standard is an examination and clinical evaluation by a physician with specialist knowledge which if stated as such should be acceptable. Ideally this should be undertaken independently by more than one specialist. Alternatively an acceptable reference standard is extended follow-up Classify as **Yes** if the test is examination and clinical evaluation by a physician with specialist knowledge and/or training, or a non-	Was there an appropriate time interval between the index test(s) and reference standard? Classify as **Yes** if the delay between the index test(s) and reference standard is considered acceptable for the majority of participants Classify as **No** if the delay between the index test(s) and reference standard is considered unacceptable for the majority of participants Classify as **Unclear** if the delay between the index test(s) and

Table 2. Indicators for the assessment of methodological quality *(Continued)*

	dom sample of individuals were recruited Classify as **Unclear** if patient selection was not clearly described		specialist with dedicated training to an acceptable standard Classify as **No** if the test result is examination and clinical evaluation by a non-specialist physician in the absence of dedicated training Classify as **Unclear** if the study does not report the experience and training of those carrying out the reference standard	reference standard is not explicitly stated
	Did the study avoid inappropriate exclusions? Classify as **Yes** if the sample consisted of apparently healthy individuals Classify as **No** if only individuals with existing PMDs were recruited Classify as **Unclear** if exclusions were not clearly described	If a threshold was used, was it pre-specified? Classify as **Yes** if the threshold was pre-specified Classify as **No** if the threshold was not pre-specified Classify as **Unclear** if it is unclear whether the threshold was pre-specified	Were the reference standard results interpreted without knowledge of the results of the index test? Classify as **Yes** if personnel clearly do not know index test results when performing the examination and clinical evaluation or evaluating follow-up data Classify as **No** if personnel clearly know index test results when performing the examination and clinical evaluation or evaluating follow-up data Classify as **Unclear** if study did not provide any information on whether personnel were blinded to the index test results	Did all patients receive the same reference standard? Classify as **Yes** if the same reference standard was used in all participants Classify as **No** if the same reference standard was not used in all participants Classify as **Unclear** if it is unclear whether different reference standards were used
		Where multiple index tests were used, were the results of the second index test interpreted without knowledge of the results of the first index test? Classify as **Yes** if index test results were interpreted without knowledge Classify as **No** if the index test results were interpreted with knowledge Classify as **Unclear** if it is unclear whether the results of the second index test were interpreted without knowledge of the results of the first index test		Were all patients included in the analysis? Classify as **Yes** if all patients were included in the analysis Classify as **No** if only some patients were included in the analysis Classify as **Unclear** if it is unclear whether all patients were included in the analysis
		Were any conflicts of interest stated?		

Table 2. Indicators for the assessment of methodological quality *(Continued)*

		Classify as **Yes** if the study declared no conflict of interest Classify as **No** if the study declared a conflict of interest Classify as **Unclear** if there was no information on conflict of interest		
Risk of bias: High/Low/Unclear	Could the selection of individuals have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct, or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns regarding applicability: High/Low/Unclear	Are there concerns that the included individuals do not match the review question?	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	
Assessment of overall risk of bias and applicability				
An overall judgement of risk of bias and applicability to the review (high, low or unclear) was undertaken based on the judgements given to each domain. If the answers to all signalling questions within a domain were judged as yes indicating low risk of bias, then the domain was judged to be at low risk of bias. A no response to a signalling question was taken as an indication of the potential for risk of bias and the authors considered this risk within the context of the study before making a decision on whether the study was a high/low risk of bias for that domain If any of the 4 domains was judged to be at high risk of bias then the study was judged to have a high risk of bias overall. If any of the 3 applicability domains was judged to be at high concern regarding applicability then the study was judged to be of high concern regarding applicability overall				

PMDs = potentially malignant disorders.

APPENDICES

Appendix 1. Search strategies for the electronic databases

Cochrane Diagnostic Test Accuracy Register search strategy

((oral* or mouth* or bucca* or "oral cavit*" or "oral mucosa" or "mouth mucosa" or lip or lips or tongue* or gingiva* or palat* or cheek* or intra-oral* or intraoral* or gum or gums or labial*) AND (tumour* or tumor* or cancer* or carcinoma* or carcinogen* or neoplas* or malignan* or metasta* or dysplas* or lesion* or ulcer* or precancer* or pre-cancer* or premalignan* or precursor* or "lichen planus" or leukoplakia or "submucous fibrosis" or "actinic keratosis" or candidiasis or erythroplakia or erythroplas* or erythroleukoplakia or hyperplas* or hyperkerato*)

Cochrane Oral Health Group's Trials Register search strategy

An updated search of the Cochrane Oral Health Group's Trials Register was conducted 30 April 2013 using the Cochrane Register of Studies software and the search strategy below:

#1 ((oral* or mouth* or bucca* or "oral cavit*" or "oral mucosa" or "mouth mucosa" or lip or lips or tongue* or gingiva* or palat* or cheek* or intra-oral* or intraoral* or gum or gums or labial*):ti,ab) AND (INREGISTER)

#2 ((tumour* or tumor* or cancer* or carcinoma* or carcinogen* or neoplas* or malignan* or metasta* or dysplas* or lesion* or ulcer* or precancer* or pre-cancer* or premalignan* or precursor* or "lichen planus" or leukoplakia or "submucous fibrosis" or "actinic keratosis" or candidiasis or erythroplakia or erythroplas* or erythroleukoplakia or hyperplas* or hyperkerato*):ti,ab) AND (INREGISTER)
#3 ((cytodiagnosis or cytophotometry or "brush biops*" or "oral cdx" or oralcdx or "modified liquid based cytology" or "exfoliat* cytolog*" or "tolonium chloride" or "toludine b*" or "toluidine b*" or tblue or t-blue or "toludine dye*" or "toludine rins*" or "toludine stain*" or "toludine wash*" or "toluidine dye*" or "toluidine rins*" or "toluidine stain*" or "toluidine wash*" or luminescence or fluorescen* or "light emitting diode*"):ti,ab) AND (INREGISTER)
#4 (((blood or saliva) AND (analys* or inspect* or test or examin*)):ti,ab) AND (INREGISTER)
#5 (("blue spectrum" or LED or luminous or "visual* adjunct*" or vizilite or microlux* or orascoptic or velscope or lumenoscope* or autofluorescen* or chemilumiescen* or spectrophotometr* or "acetic acid" or acetowhite or "tumor marker*" or "tumour marker*" or "neoplas* marker*"):ti,ab) AND (INREGISTER)
#6 ((diagnos* AND (exam* or histolog* or check* or screen*)):ti,ab) AND (INREGISTER)
#7 (#1 and #2) AND (INREGISTER)
#8 (#3 or #4 or #5 or #6) AND (INREGISTER)
#9 (#7 and #8) AND (INREGISTER)

A previous search was conducted in June 2011 using the Procite software and the search strategies below:

((oral* or mouth* or bucca* or "oral cavit*" or "oral mucosa" or "mouth mucosa" or lip or lips or tongue* or gingiva* or palat* or cheek* or intra-oral* or intraoral* or gum or gums or labial*) AND (tumour* or tumor* or cancer* or carcinoma* or carcinogen* or neoplas* or malignan* or metasta* or dysplas* or lesion* or ulcer* or precancer* or pre-cancer* or premalignan* or precursor* or "lichen planus" or leukoplakia or "submucous fibrosis" or "actinic keratosis" or candidiasis or erythroplakia or erythroplas* or erythroleukoplakia or hyperplas* or hyperkerato*) AND (cytodiagnosis or cytophotometry or "brush biops*" or "oral cdx" or oralcdx or "modified liquid based cytology" or "exfoliat* cytolog*" or "tolonium chloride" or "toludine b*" or "toluidine b*" or tblue or t-blue or "toludine dye*" or "toludine rins*" or "toludine stain*" or "toludine wash*" or "toluidine dye*" or "toluidine rins*" or "toluidine stain*" or "toluidine wash*" or luminescence or fluorescen* or "light emitting diode*"))

((oral* or mouth* or bucca* or "oral cavit*" or "oral mucosa" or "mouth mucosa" or lip or lips or tongue* or gingiva* or palat* or cheek* or intra-oral* or intraoral* or gum or gums or labial*) AND (tumour* or tumor* or cancer* or carcinoma* or carcinogen* or neoplas* or malignan* or metasta* or dysplas* or lesion* or ulcer* or precancer* or pre-cancer* or premalignan* or precursor* or "lichen planus" or leukoplakia or "submucous fibrosis" or "actinic keratosis" or candidiasis or erythroplakia or erythroplas* or erythroleukoplakia or hyperplas* or hyperkerato*) AND ("blue spectrum" or LED or luminous or "visual* adjunct*" or vizilite or microlux* or orascoptic or velscope or lumenoscope* or autofluorescen* or chemilumiescen* or spectrophotometr* or "acetic acid" or acetowhite or "tumor marker*" or "tumour marker*" or "neoplas* marker*"))

((oral* or mouth* or bucca* or "oral cavit*" or "oral mucosa" or "mouth mucosa" or lip or lips or tongue* or gingiva* or palat* or cheek* or intra-oral* or intraoral* or gum or gums or labial*) AND (tumour* or tumor* or cancer* or carcinoma* or carcinogen* or neoplas* or malignan* or metasta* or dysplas* or lesion* or ulcer* or precancer* or pre-cancer* or premalignan* or precursor* or "lichen planus" or leukoplakia or "submucous fibrosis" or "actinic keratosis" or candidiasis or erythroplakia or erythroplas* or erythroleukoplakia or hyperplas* or hyperkerato*) AND (diagno* and (blood or saliva) and (analys* or inspect* or test* or examin*)))

((oral* or mouth* or bucca* or "oral cavit*" or "oral mucosa" or "mouth mucosa" or lip or lips or tongue* or gingiva* or palat* or cheek* or intra-oral* or intraoral* or gum or gums or labial*) AND (tumour* or tumor* or cancer* or carcinoma* or carcinogen* or neoplas* or malignan* or metasta* or dysplas* or lesion* or ulcer* or precancer* or pre-cancer* or premalignan* or precursor* or "lichen planus" or leukoplakia or "submucous fibrosis" or "actinic keratosis" or candidiasis or erythroplakia or erythroplas* or erythroleukoplakia or hyperplas* or hyperkerato*) AND (diagnos* AND (exam* or histolog* or check or inspect* or screen*)))

MEDLINE via OVID search strategy

1. exp Mouth/
2. Cheek/
3. or/1-2
4. exp Carcinoma, squamous cell/di
5. exp Precancerous conditions/di
6. (tumor$ or tumour$ or cancer$ or carcinoma$ or carcinogen$ or neoplas$ or malignan$ or metasta$ or dysplas$ or lesion$ or ulcer $).tw,ot.
7. (pre-cancer$ or precancer$ or premalignan$ or precursor$ or "lichen planus" or leukoplakia or "submucous fibrosis" or (actinic adj2 keratosis) or candidiasis or erythroplakia or erythroplas$ or erythroleukoplakia or hyperplas$ or hyperkeratos$).tw,ot.
8. or/4-7
9. 3 and 8
10. exp Mouth neoplasms/di
11. Lichen Planus, Oral/di
12. Oral submucous fibrosis/di

13. Oral candidiasis/di
14. ((oral$ or mouth$ or bucca$ or "oral cavit$" or (oral adj mucosa$) or (mouth adj mucosa$) or lip or lips or tongue$ or gingiv$ or palat$ or cheek$ or "intra oral$" or intraoral$ or gum or gums or labial$) adj3 (tumor$ or tumour$ or cancer$ or carcinoma$ or carcinogen$ or neoplas$ or malignan$ or metasta$ or dysplas$ or lesion$ or ulcer$ or pre-cancer$ or precancer$ or premalignan$ or precursor$ or "lichen planus" or leukoplakia or "submucous fibrosis" or (actinic adj2 keratosis) or candidiasis or erythroplakia or erythroplas$ or erythroleukoplakia or hyperplas$ or hyperkerato$)).tw,ot.
15. or/10-14
16. 9 or 15
17. Cytodiagnosis/
18. Cytological techniques/
19. Cytophotometry/
20. (brush adj3 biops$).tw,ot.
21. ("oral cdx" or oralcdx).tw,ot.
22. ("modified liquid based cytology" or (exfoliat$ adj3 cytolog$)).tw,ot.
23. (brush$ and (cytodiagnosis or cytopathology)).tw,ot.
24. Tolonium chloride/du
25. Coloring agents/du
26. ("tolonium chloride" or "tolu?dine blue" or "tolu?dine b" or tblue or t-blue).tw,ot.
27. (tolu?dine adj6 (dye$ or rins$ or stain$ or wash$)).tw,ot.
28. exp Luminescence/du
29. Fluorescence/
30. Spectrometry, fluorescence/
31. exp Luminescent Agents/du
32. Light/du
33. Tomography, Optical Coherence/
34. (visual$ adj5 ("light emitting diode" or "blue spectrum" or LED or luminous$)).tw,ot.
35. (visuali?ation adj3 adjunct$).tw,ot.
36. (vizilite or microlux$ or orascoptic or velscope).tw,ot.
37. lumenoscop$.tw,ot.
38. ((tumor$ or tumour$ or cancer$ or carcinoma$ or neoplas$ or carcinogen$ or malignan$ or metata$ or lesion$ or ulcer$) adj5 (fluorescen$ or autofluorescen$ or luminescen$ or chemiluminescen$)).tw,ot.
39. (tissue adj3 reflect$).tw,ot.
40. Spectrophotometry/
41. Acetic acid/du
42. (acetic acid adj3 (wash$ or rins$)).tw,ot.
43. acetowhite.tw,ot.
44. Saliva/an, ch
45. Tumor Markers, Biological/an
46. (("tumo?r marker$" or "neoplas$ marker$") adj3 (blood or saliva)).tw,ot.
47. ((analy$ or screen$ or test$ or examin$) adj3 (blood or saliva)).tw,ot.
48. Diagnosis, Oral/
49. Mass screening/
50. Physical examination/
51. ((oral$ or mouth$) adj5 (exam$ or histolog$ or check$ or inspect$)).tw,ot.
52. (visual$ adj3 (exam$ or inspect$ or screen$)).tw,ot.
53. or/17-52
54. 16 and 53

EMBASE via OVID search strategy

1. exp Mouth/
2. Cheek/
3. or/1-2
4. exp Squamous cell carcinoma/di
5. exp Precancer/di
6. (tumor$ or tumour$ or cancer$ or carcinoma$ or carcinogen$ or neoplas$ or malignan$ or metasta$ or dysplas$ or lesion$ or ulcer$).tw,ot.
7. (pre-cancer$ or precancer$ or premalignan$ or precursor$ or "lichen planus" or leukoplakia or "submucous fibrosis" or (actinic adj2 keratosis) or candidiasis or erythroplakia or erythroplas$ or erythroleukoplakia or hyperplas$ or hyperkeratos$).tw,ot.
8. or/4-7
9. 3 and 8
10. exp Mouth tumor/di

11. Lichen planus/di
12. Thrush/di
13. ((oral$ or mouth$ or bucca$ or "oral cavit$" or (oral adj mucosa$) or (mouth adj mucosa$) or lip or lips or tongue$ or gingiv$ or palat$ or cheek$ or "intra oral$" or intraoral$ or gum or gums or labial$) adj3 (tumor$ or tumour$ or cancer$ or carcinoma$ or carcinogen$ or neoplas$ or malignan$ or metasta$ or dysplas$ or lesion$ or ulcer$ or pre-cancer$ or precancer$ or premalignan$ or precursor$ or "lichen planus" or leukoplakia or "submucous fibrosis" or (actinic adj2 keratosis) or candidiasis or erythroplakia or erythroplas$ or erythroleukoplakia or hyperplas$ or hyperkerato$)).tw,ot.
14. or/10-13
15. 9 or 14
16. Cancer cytodiagnosis/
17. Cytophotometry/
18. (brush adj3 biops$).tw,ot.
19. ("oral cdx" or oralcdx).tw,ot.
20. ("modified liquid based cytology" or (exfoliat$ adj3 cytolog$)).tw,ot.
21. (brush$ and (cytodiagnosis or cytopathology)).tw,ot.
22. Tolonium chloride/
23. Coloring agent/
24. ("tolonium chloride" or "tolu?dine blue" or "tolu?dine b" or tblue or t-blue).tw,ot.
25. (tolu?dine adj6 (dye$ or rins$ or stain$ or wash$)).tw,ot.
26. exp Luminescence/
27. Fluorescence/
28. Spectrofluorometry/
29. exp Luminescent Agents/
30. Light/
31. Tomography, Optical Coherence/
32. (visual$ adj5 ("light emitting diode" or "blue spectrum" or LED or luminous$)).tw,ot.
33. (visuali?ation adj3 adjunct$).tw,ot.
34. (vizilite or microlux$ or orascoptic or velscope).tw,ot.
35. lumenoscop$.tw,ot.
36. ((tumor$ or tumour$ or cancer$ or carcinoma$ or neoplas$ or carcinogen$ or malignan$ or metata$ or lesion$ or ulcer$) adj5 (fluorescen$ or autofluorescen$ or luminescen$ or chemiluminescen$)).tw,ot.
37. (tissue adj3 reflect$).tw,ot.
38. Spectrophotometry/
39. Acetic acid/
40. (acetic acid adj3 (wash$ or rins$)).tw,ot.
41. acetowhite.tw,ot.
42. Saliva/
43. Tumor Marker/
44. (("tumo?r marker$" or "neoplas$ marker$") adj3 (blood or saliva)).tw,ot.
45. ((analy$ or screen$ or test$ or examin$) adj3 (blood or saliva)).tw,ot.
46. Mass screening/
47. Physical examination/
48. ((oral$ or mouth$) adj5 (diagnos$ or exam$ or histolog$ or check$ or inspect$)).tw,ot.
49. (visual$ adj3 (exam$ or inspect$ or screen$)).tw,ot.
50. or/16-49
51. 15 and 50

MEDION search strategy

Searched using the code C (malignancies), and screened the results for oral cancer terms.

CONTRIBUTIONS OF AUTHORS

Tanya Walsh wrote this review with contributions from Joseph Liu, Paul Brocklehurst, Anne-Marie Glenny, Graham Ogden, Crispian Scully, Ross Kerr, Mark Lingen and Saman Warnakulasuriya.

DECLARATIONS OF INTEREST

None declared at this point in time.

SOURCES OF SUPPORT

Internal sources

- MAHSC, UK.

  The Cochrane Oral Health Group is supported by the Manchester Academic Health Sciences Centre (MAHSC) and the NIHR Manchester Biomedical Research Centre.
- NIHR-MBRC, UK.

  Health Sciences Research, School of Dentistry is supported by the Manchester Academic Health Sciences Centre (MAHSC) and the NIHR Manchester Biomedical Research Centre.
- SCDEP, UK.

  This review has been supported by the Scottish Dental Clinical Effectiveness Programme.

External sources

- Cochrane Oral Health Group Global Alliance, UK.

  All reviews in the Cochrane Oral Health Group are supported by Global Alliance member organisations (British Association of Oral Surgeons, UK; British Orthodontic Society, UK; British Society of Paediatric Dentistry, UK; British Society of Periodontology, UK; Canadian Dental Hygienists Association, Canada; National Center for Dental Hygiene Research & Practice, USA; Mayo Clinic, USA; New York University College of Dentistry, USA; and Royal College of Surgeons of Edinburgh, UK) providing funding for the editorial process (http://ohg.cochrane.org/).
- National Institute for Health Research (NIHR), UK.

  CRG funding acknowledgement:
  The NIHR is the largest single funder of the Cochrane Oral Health Group.

  Disclaimer:
  The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the NIHR, NHS or the Department of Health.
- Centers for Disease Control and Prevention (CDC), USA.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We have included mouth self examination as an additional index test in this review.

We have removed the index test training and calibration signalling question from the QUADAS -2 assessment of methodological quality. The diversity of index tests meant we were unable to uniformly apply this criterion to all the studies. For example for the conventional oral examination, the index test was conducted by a variety of personnel of differing clinical experience. Where we would expect that, for example basic health workers would need specific training and the adequacy of the training would be evaluated, the same cannot be said for experienced general dental practitioners or oral specialists. The challenge is even greater when considering different index tests; for example training and calibration of mouth self examination. For all index tests, we would expect that any training given would be reported and any diagnostic criteria followed in the index test assessment would have been piloted/validated. All study information pertaining to how the index test was carried out and interpreted is detailed in the Characteristics of included studies tables.

INDEX TERMS

Medical Subject Headings (MeSH)

*Health Status; Early Detection of Cancer [methods] [*standards]; Lip Neoplasms [diagnosis]; Mouth Neoplasms [*diagnosis]; Randomized Controlled Trials as Topic; Sensitivity and Specificity

MeSH check words

Adult; Humans