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

Walsh, Tanya; Liu, Joseph L. Y.; Brocklehurst, Paul; Lingen, Mark; Kerr, Alexander R.; Ogden, Graham; Warnakulasuriya, Saman; Scully, Crispian

Published in:
Cochrane Database of Systematic Reviews

DOI:
10.1002/14651858.CD010173

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Discovery Research Portal

Citation for published version (APA):
Clinical assessment to screen for the detection of oral cavity cancer and potentially malignant disorders in apparently healthy adults (Protocol)


This is a reprint of a Cochrane protocol, prepared and maintained by The Cochrane Collaboration and published in The Cochrane Library 2012, Issue 11

http://www.thecochranelibrary.com

Copyright © 2012 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.
# Table of Contents

- **Header** .......................................................... 1
- **Abstract** ......................................................... 1
- **Background** ...................................................... 2
- **Objectives** ....................................................... 4
- **Methods** .......................................................... 5
- **Acknowledgements** ............................................... 7
- **References** ....................................................... 8
- **Additional Tables** ................................................. 10
- **Appendices** ...................................................... 15
- **History** ............................................................ 16
- **Contributions of Authors** ....................................... 17
- **Declarations of Interest** ......................................... 17
- **Sources of Support** ............................................... 17
Clinical assessment to screen for the detection of oral cavity cancer and potentially malignant disorders in apparently healthy adults

Tanya Walsh1, Joseph LY Liu2, Paul Brocklehurst1, Mark Lingen3, Alexander R Kerr4, Graham Ogden5, Saman Warnakulasuriya6, Crispian Scully7

1School of Dentistry, The University of Manchester, Manchester, UK. 2University of Dundee, Dental Health Services Research Unit, Scottish Dental Clinical Effectiveness Programme, NHS Education for Scotland, Dundee, UK. 3Pritzker School of Medicine, Division of Biological Sciences, Department of Pathology, University of Chicago, Chicago, Illinois, USA. 4Department of Oral and Maxillofacial Pathology, Radiology and Medicine, New York University College of Dentistry, New York, USA. 5Unit of Oral Surgery & Medicine, School of Dentistry, University of Dundee, Dundee, UK. 6Department of Oral Medicine, King’s College London, London, UK. 7Eastman Dental Institute for Oral Health Care Sciences, 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.


Copyright © 2012 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

This is the protocol for a review and there is no abstract. The objectives are as follows:

The objective of this review is to estimate the accuracy of the conventional oral examination (COE) used singly or in combination with another index test (Additional Table 1) as a screening test for the detection of oral cancer and potentially malignant disorders (PMD) of the lip and oral cavity of apparently healthy adults.

The secondary objective of this review is to estimate the accuracy of the different tests with COE, when compared with each other.

We will use meta-regression to explore possible sources of heterogeneity, ways in which the observed diagnostic test accuracy varies according to particular characteristics. Covariates to be included in these analyses will include:

- Characteristics of the study sample: prevalence of carcinoma or PMD disease 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).
- Attributes of the screening programme: prospective organised or opportunistic, type of reference standard (examination and clinical evaluation by physician with specialist knowledge or extended follow-up), operator (dental or general practice professionals or other healthcare workers).
**BACKGROUND**

**Target condition being diagnosed**

The target conditions of interest are oral squamous cell carcinoma (OSCC), the most common form of oral cavity cancer (Scully 2000a), and potentially malignant disorders (PMD), of the lip and oral cavity in apparently healthy adults. At a meeting of international oral cancer and precancer experts held in 2005, the concept of precancer, along with issues surrounding classification and definition, aetiology, diagnosis and management was extensively discussed. Through consensus, the term 'potentially malignant disorders' was selected to convey the fact that not all precancerous lesions and conditions will transform to cancer, but rather that there is the potential for malignant transformation (van der Waal 2009; Warnakulasuriya 2007).

The natural history of oral cancer is not fully understood, given variations in disease processes and dysplastic changes in PMD (Napier 2008; Scully 2009). Most oral carcinomas are preceded by PMD of which erythroplakia, non-homogeneous leukoplakia, erosive lichen planus, oral submucous fibrosis and actinic keratosis are perhaps the most important (Warnakulasuriya 2007). The concept of a two-step process of cancer development of the oral mucosa (i.e. precursor to established lesion) is established. Oral leukoplakia is the best-known precursor lesion and between <1% and 18% develop into oral cancer. The original 1978 World Health Organization (WHO) definition of oral leukoplakia has been revised to read “The term leukoplakia should be used to recognize white plaques of questionable risk having excluded (other) known diseases or disorders that carry no increased risk for cancer” (Warnakulasuriya 2007).

The presence of epithelial dysplasia can help predict malignant development but the diagnosis is essentially subjective, with not all lesions exhibiting dysplasia, some becoming malignant and some regressing. Carcinoma can also develop from lesions in which epithelial dysplasia was not previously diagnosed. Numerous attempts have been made to relate biological characteristics to the malignant potential of leukoplakias, but finding a definitive characteristic remains elusive (Reibel 2003). Most authorities regard leukoplakia as a dynamic rather than a static process, in terms of its progression and the development of malignancy (Napier 2008). For example, Jaber et al (Jaber 2003) followed up 630 patients who attended oral medicine clinics in London and Bristol between 1972 and 1996. The majority of oral leukoplakia (43.8%) had features of mild dysplasia, 30% moderate and 24.7% severe oral epithelial dysplasia. Whilst this is strongly suggestive that dysplasia is typically associated with oral leukoplakia, it can also be seen in patients with reactive epithelial changes such as candidosis, viral infections, lichen planus, and denture-induced hyperplasia. The study concluded that predicting the severity of histopathological change from clinical examination alone remains difficult. 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 (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 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 5 years. For example, the incidence of 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 excision of some PMD can prevent malignancy, or if malignancy is detected, there is some evidence that appropriate treatment can reduce disease severity and improve survival rates (Brocklehurst 2010b; van der Waal 2009; Warnakulasuriya 2007). Leukoplakias can be treated by a number of methods. According to Lodi et al’s systematic review (Lodi 2006), the effectiveness of surgical interventions, including laser therapy and cryotherapy, has not been studied by means of a randomised controlled trial (RCT) with a no treatment/placebo arm. Vitamin A and retinoids have been tested by five RCTs, two studies investigated beta carotene or carotenoids, the other drugs tested were bleomycin (one study), mixed tea (one study) and ketorolac (one study). 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. Where resolution of a lesion is observed, relapses and adverse effects are common.

Technologies to treat and manage oral cancer have progressed substantially, as shown by systematic reviews of RCT of interventions, e.g. Bessell 2011; Furness 2011; Glenny 2010. Once progressed to frank malignancy, traditional treatment is 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 radiotherapy and chemotherapy, adversely affecting an individual’s quality of life. The 5-year survival following diagnosis has remained at around 50% for the past 30 years in most countries (Parkin 2001; Warnakulasuriya 2009). 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)**

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

Copyright © 2012 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.
The standard screening of apparently healthy adults for oral cancer and PMD is a systematic and thorough visual inspection and examination of the oral mucosa and palpation of the neck under normal (incandescent) light. In most instances this is carried out by a frontline clinician such as a general dentist. This conventional visual and tactile oral examination (COE) used as a screen 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 of further investigation is variable nationally or internationally, with different investigational pathways. For instance this could take the form of examination by an oral medicine specialist or specialist oral surgeon at a specialist clinic or hospital facility.

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 COE to improve earlier detection of oral cavity cancer and PMD (Fedele 2009; Leston 2010; Lingen 2008; Patton 2008; Rethman 2010). These include:

- vital staining (Toluidine Blue, Tolonium chloride)
- light-based detection (such as ViziLite and ViziLite Plus, Microlus/DL, VEILscope, Orascopic DK, Identafi 3000)
- blood and saliva analysis.

Vital staining 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. Blood analysis and saliva analysis are more novel tests at an early stage of evaluation. Of the index tests listed above, vital staining, light-based detection and blood and saliva analysis could be used as screening adjuncts to the COE (Additional Table 1). In this review, we will restrict vital staining index tests to those applied in a rinse form. A companion review (Diagnostic tests for oral cancer and potentially malignant disorders in patients presenting with clinically evident lesions) will restrict vital staining index tests to those applied to a lesion that has been visualized. Where access to general dental practitioners or general 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.

It is worth noting there are regional differences in regulations on the use of some of the above tests. For example, Toluidine Blue having been consistently rejected as a stand alone technique, is not cleared to be a stand alone screening technique in the United States. It is included in the ViziLite Plus system. However, the Toluidine Blue only component is approved by the US Food and Drug Administration (FDA) as a marking device.

Of the index tests listed above, all have the potential to be used as adjuncts to the COE (Additional Table 1) by healthcare workers or clinicians undertaking screening in 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. 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.

**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 developing countries as in developed 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). The human papillomavirus (HPV) is also an important risk factor in oral squamous cell carcinoma and there is some evidence that HPV positive cancers associate with better survival (Ang 2010). 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 5 males to 1 female diagnosed with oral cancers in the 1960s to less than 2 to 1 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). The 5-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 cancer or PMD is highly dependent on whether ‘at risk’ individuals present for screening examination. Early diagnosis 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), 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 developed countries, although opportunistic screening has been advocated (Brocklehurst 2010a). 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. A province-wide programme is being evaluated in British Columbia, Canada (Rosin 2006) but the evaluation is ongoing and no final results have been reported to date. Brocklehurst et al’s systematic review identified only one RCT using visual examination with a follow-up period of 9 years which was carried out in India. The authors of the review concluded that the current evidence is insufficient to recommend population-based screening and suggested that opportunistic screening of high risk groups may potentially improve outcomes (Brocklehurst 2010a).

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, the results of the only completed randomised controlled trial undertaken on oral cancer screening as an intervention demonstrated that the yield from both whole population and opportunistic approaches is likely to be important (Brocklehurst 2010a). Many individuals with risk factors may not attend the dentist and are therefore not amenable to an opportunistic approach (Netuveli 2006; Yusof 2006). Furthermore, the evidence suggests that diagnostic/screening tools based on machine learning such as Speight et al’s (Speight 2006) can provide useful but preliminary evidence, as they tend to be limited by “the fitting of models that are implausible and the tendency . . . to understate misclassification errors” (Liu 2006).

Reviews assessing the test accuracy of a conventional oral examination as a 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 aim to identify screening tests for oral cancer and PMD 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 apparently healthy 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 will include both prospective investigations of organised screening programmes and prospective opportunistic screening. It is important that this review consider both, as opportunistic screening of patients attending in a general practice setting 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 PMDs 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; Furrnness 2011; Glenny 2010) and screening programmes for the early detection and prevention of oral cancer (Brocklehurst 2010a). This screening test accuracy review will complement the intervention reviews.

**OBJECTIVES**

The objective of this review is to estimate the accuracy of the conventional oral examination (COE) used singly or in combination with another index test (Additional Table 1) as a screening test for the detection of oral cancer and potentially malignant disorders (PMD) of the lip and oral cavity of apparently healthy adults.

**Secondary objectives**

The secondary objective of this review is to estimate the accuracy of the different tests with COE, when compared with each other.

**Investigation of sources of heterogeneity**

We will use meta-regression to explore possible sources of heterogeneity, ways in which the observed diagnostic test accuracy varies according to particular characteristics. Covariates to be included in these analyses will include.

- Characteristics of the study sample: prevalence of carcinoma or PMD disease 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).
- Attributes of the screening programme: prospective organised or opportunistic, type of reference standard (examination and clinical evaluation by physician with specialist...
knowledge or extended follow-up), operator (dental or general practice professionals or other healthcare workers).

M E T H O D S

Criteria for considering studies for this review

Types of studies
Studies of clinical cohorts of apparently healthy adults which report the diagnostic accuracy of the conventional oral examination (COE) used singly or in combination with an index test listed in Additional Table 1, for oral cancer and potentially malignant disorders (PMD) with respect to one of the reference standards. These will include cross-sectional studies (or consecutive series) and randomised controlled trials (RCTs) of test accuracy. We will exclude case series and case control studies which can lead to inflated estimates of prevalence and test accuracy (Whiting 2004). We will exclude studies reported in abstract form alone, uncontrolled reports and randomised controlled trials of intervention studies. Where randomised or paired comparative designs are available these will be included in the review and analysed separately. Studies should report data for true positives, false positives, true negatives and false negatives for each test. Only studies reporting results at an individual level (as opposed to a lesion level) will be included.

Participants
Apparently healthy adults (aged 16 years or over) attending an organised screening programme or screened during attendance at a dental or general practice examination.

Index tests
The COE used as a screen, alone or in combination with any other screening tests previously listed (Additional Table 1). The COE (comparator test) is the initial point of the screen, which all adults will receive. The index test will be used as an adjunct following the COE irrespective of whether oral cancer or PMD is 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).

Comparator tests
The comparator test is the COE used as a stand alone screen as previously described. This is the standard screen commonly used in clinical practice, although problems with its use have been identified (Lingen 2008).

Target conditions
Following the consensus views of the expert working group of the WHO Collaborating Centre for Oral Cancer and Precancer, the target conditions of the lip or oral cavity of interest are noted as:
- Carcinoma
  - Squamous cell carcinoma
- 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 is examination and clinical evaluation by a physician with specialist knowledge, such as an oral and maxillofacial pathologist or oral medicine specialist, working to the current diagnostic guidelines of their locality. We will include the diagnostic protocol / guidelines used in each study in the ‘Characteristics of included studies’ table. Studies with confirmatory biopsy of individuals who screened negative by the index test may exist, but are questionable ethically (Downer 2004). Confirmation of individuals screened negative by an index test may be done by extended follow-up of a minimum of 5 years.

Search methods for identification of studies

Electronic searches
The following databases will be searched using a highly sensitive search strategy:
- Cochrane Oral Health Group’s Trials Register (to present)
- Cochrane Register of Diagnostic Test Accuracy Studies (to present)
- MEDLINE (1948 to present)
- EMBASE (1980 to present)
- MEDION (2003 to present).

The MEDLINE search strategy outlined in Appendix 1 will be modified for the listed databases. The search will not be limited by language or publication status. Non-English articles will be translated, unless a translator cannot be found through The Cochrane Collaboration.

The search strategy has been constructed in accordance with this protocol and that of a companion Cochrane Diagnostic Test Accuracy review 'Diagnostic tests for oral cancer in patients presenting..."
with clinically suspicious lesions’ being undertaken concurrently by the same review team.

**Searching other resources**

We will also search relevant conference proceedings. We will 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**

Titles and abstracts of all articles identified from the searches will be independently assessed by two review authors (Tanya Walsh (TW) and Joseph Liu (JL)). For articles that appear to meet the inclusion criteria, or where a clear decision cannot be made from scanning the title and abstract alone, full reports will be obtained. Two review authors (TW and JL) will independently assess each report for inclusion. Where disagreements occur, the review authors will attempt to resolve these by discussion. If needed, a third review author (Paul Brocklehurst (PB)) will be asked to help resolve the discrepancies in consultation with the other two review authors.

**Data extraction and management**

Two review authors (TW and JL) will independently extract data using a piloted data collection form. Discrepancies will be resolved through discussion. If an agreement cannot be reached, a third review author will be consulted (PB). Study authors will be contacted to obtain relevant missing data if these are not available in the printed report.

The following data will be recorded from each study.

- Sample characteristics (age, sex, socio-economic status, risk factors (e.g. HPV status, prevalence of tobacco use and alcohol consumption), number of participants / lesions, lesion site)
- Setting (country, disease prevalence, type of screening)
- The type of index test(s) used (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).

**Assessment of methodological quality**

The revised QUADAS tool, QUADAS-2 (Whiting 2011) will be used to assess the quality of the primary diagnostic studies over four key domains: patient selection, index test, reference standard and flow and timing of participants through the study.

In the first phase of the tool, the review question will be stated in terms of patient sample, index test, reference standard and target condition. This information is detailed in the Criteria for considering studies for this review section of this protocol. In phase two, the QUADAS-2 tool will be tailored to use with this review (Additional Table 2). Review specific guidance will be used to facilitate documentation of the pertinent descriptive information contained in the primary studies. Customised instructions to aid judgement of the signalling questions will be given (following Patton 2008). Two core signalling items 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). Three additional signalling items relating to commercial funding, training and calibration and multiple index tests have been added to the core signalling questions.

In phase three, a flow diagram will be drawn. In the final phase, an overall judgement of risk of bias and applicability is to be undertaken. A risk of bias judgement (‘high’, ‘low’ or ‘unclear’) will be reached for each domain. If the answers to all signalling questions within a domain are judged as ‘yes’ indicating low risk of bias, then the domain will be judged to be at low risk of bias. If any signalling question within a domain is judged as ‘no’ indicating high risk of bias then this indicates that potential bias exists. This will be followed by a judgement for concerns regarding applicability for the patient selection, index test and reference standard domains.

We will pilot the use of the QUADAS-2 tool independently on five study reports. Where disagreements occur between the two review authors the review specific descriptions will be clarified until consistency is obtained.

Results of the quality assessment for all included studies will be summarised in a narrative report. A summary tabular presentation of the results for each domain will be also provided separately for risk of bias and concerns regarding applicability, along with a graphical display summarising this information.

**Statistical analysis and data synthesis**

Data for the true positive, true negative, false positive and false negative values for each test in each study will be tabulated. For each index test, estimates of diagnostic accuracy as sensitivity and specificity along with their 95% confidence intervals will be displayed as coupled forest plots, and plotted in receiver operating characteristic (ROC) space.

Meta-analysis will be used to combine the results of studies for each index test. Random-effects models will be used. If however there are too few studies to reliably estimate between study variability then fixed-effect models will be used. The statistical software SAS will be used throughout (SAS Institute Inc., Cary, USA).

Classification of responses of the various tests are given in Additional Table 1. Where a common threshold of response is reported, the analysis will estimate the expected values of sensitivity and specificity (Bivariate approach Reitsma 2005). Where there is a variation in thresholds, the expected hierarchical summary ROC
curve for the tests across different thresholds will be estimated (Macaskill 2010; Rutter 2001). Hierarchical SROC curves will be fitted using the Proc NLMixed procedure in SAS. The proposed analysis is subject to change based on information reported in the primary studies. For example, if there is little variation in the positive thresholds of the blood and salivary index tests, it will not be appropriate to attempt to fit a summary ROC curve (Macaskill 2010). Where the primary studies have published more than one threshold result, accuracy estimates will be reported for all the thresholds. Data from all thresholds will be extracted and used in the analysis.

The statistical analysis plan can be specified as follows.

Primary analyses: The primary analyses will compare the COE used singly or in combination with any other test (Additional Table 1) with the reference standard.

- COE
- COE with vital rinse
- COE with light detection
- COE with blood / salivary analysis.

This will either estimate the average sensitivity and specificity of a test or describe the variation in sensitivity and specificity at different thresholds by estimating a hierarchical summary ROC curve, depending on nature of the index tests. Parameter estimates will include sensitivity, specificity and their correlation or hierarchical summary ROC curve.

Secondary analysis: The comparative accuracy of the index tests with the reference standard will be the focus of the secondary analyses. A preliminary analysis will graphically display the sensitivities and specificities of the index tests. This will be followed by a series of indirect pairwise analyses and structured as follows.

- COE versus COE with vital rinse
- COE versus COE with light detection
- COE versus COE with blood / salivary analysis
- COE with vital rinse versus COE with light detection
- COE with vital rinse versus COE with blood / salivary analysis
- COE with light detection versus COE with blood / salivary analysis.

All studies will be included in each pairwise comparison. Where studies of direct comparisons exist (i.e. paired data from all patients or randomising individuals to different tests) the results of these studies will be analysed and reported separately.

The methodology used is akin to the investigation of heterogeneity (as below) i.e. adding a covariate for test type into the Bivariate or Hierarchical SROC analysis.

Investigations of heterogeneity

Meta-regression analyses will be carried out to explore possible sources of heterogeneity. Covariates to be included in these analyses will include 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 practice professionals or other healthcare workers). Different thresholds of binary categorisations will also be considered.

The log likelihood of models including the covariate will be compared to those models without the covariate. Formal model comparisons of either the Hierarchical SROC or Bivariate models will be undertaken using the Likelihood Ratio statistic to statistically compare the effects of adding or removing covariates. If statistical evidence of heterogeneity is found further investigations will be undertaken.

Sensitivity analyses

Sensitivity analyses will be conducted. This will entail restricting the analysis to studies where the reference standard is examination and clinical evaluation by physician with specialist knowledge or where differential verification is avoided.

Assessment of reporting bias

Tests for reporting bias will not be conducted because current tests are misleading when applied to systematic reviews of diagnostic test accuracy (Leeflang 2008; Tang 2000).

Acknowledgements

We would like to thank NHS Education Scotland and the Scottish Dental Clinical Effectiveness Programme for their support on this review; Anne Littlewood for her advice on searching the literature; Derek Richards for his thoughtful feedback on the protocol; and Luisa Fernandez Mauleffinch and Philip Riley for their assistance in facilitating this review.
REFERENCES

Additional references

American Cancer Society 1992

Ang 2010

Bessel 2011

Brocklehurst 2010a

Brocklehurst 2010b

Buchen 2011

Cancer Research UK

Conway 2008

Downer 2004

Faggiano 1997

Fedele 2009

Ferlay 2010

Freedman 2007

Furness 2011

Garg 2005

Glenny 2010

Jaber 2003

La Vecchia 1997

Landis 1977

Lee 2009

Leeflang 2008
Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM, Cochrane Diagnostic Test Accuracy Working Group. Systematic

**Leeflang 2008a**


**Leston 2010**


**Li 2006**


**Lingen 2008**


**Liu 2006**


**Lodi 2006**


**Macaskill 2010**


**Macfarlane 1995**


**McGurk 2010**


**Moles 2002**


**Napier 2008**


**Netuveli 2006**


**Ogden 2005**


**Parkin 2001**


**Petti 2003**


**Reibel 2003**


**Reitsma 2005**


**Rethman 2010**


**Rogers 2009**


**Rosin 2006**

Rutter 2001

Sankaranarayanan 1997

Scully 2000a

Scully 2000b

Scully 2009

Silverman 1984

Speight 2006

Tang 2000

US Preventive Services Task Force

van der Waal 2009

Warnakulasuriya 2007

Warnakulasuriya 2009

Whiting 2004

Whiting 2011

Wyatt 1995

Yusof 2006

* Indicates the major publication for the study

ADDITIONAL TABLES

Table 1. Screening tests for PMDs and oral cavity cancer

<table>
<thead>
<tr>
<th>Test</th>
<th>Characteristics</th>
<th>Classification of response</th>
<th>Other information</th>
</tr>
</thead>
</table>
| 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 test result; the absence of any oral mucosal abnormalities is classified as a negative test result | Traditionally been used as an oral cancer screen, but its utility is debated (Lingen 2008). **Advantages:** quick and easy once trained, minimally invasive. **Disadvantages:** oral mucosal abnormalities are not necessarily clinically or biologically ma-
Table 1. Screening tests for PMDs and oral cavity cancer  
(Continued)

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Description</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vital rinsing (e.g. Toluidine Blue, Tolonium chloride)</td>
<td>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.</td>
<td><strong>Advantages</strong>: ability to define areas that could be malignant or abnormal but cannot be seen; assess the extent of the PMD for excision.</td>
<td><strong>Disadvantages</strong>: 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.</td>
</tr>
<tr>
<td>Light-based detection (e.g. ViziLite and ViziLite plus, Microlux/DL, VELscope, Identafi 3000)</td>
<td>Light-based systems to identify premalignant and malignant lesions and to highlight their presence through tissue reflectance (Leston 2010; Lingen 2008; Patton 2008). E.g. using ViziLite Plus or Microlux/DL, the procedure is as follows (Lingen 2008).</td>
<td><strong>Advantages</strong>: simple to use; non-invasive; do not require consumable reagents; provide real time results; can be performed by a wide range of operators after a short training period.</td>
<td><strong>Disadvantages</strong>: 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.</td>
</tr>
</tbody>
</table>
| Blood and saliva analysis. | These novel technologies are at an early stage of development and evaluation. Analysis of blood or saliva samples which tests for the presence of Cut-off probabilities vary widely and are dependent on the individual biomarker or combination of biomarkers examined. | **Advantages**: non-invasive (saliva tests) or minimally invasive (blood tests). | **Disadvantages**: there is a tendency for the estimated diag-
bio-markers of PMD and oral cancer (Brinkmann 2011; Lee 2009; Li 2006). Molecular markers for diagnosis include changes in cellular DNA, altered mRNA transcripts, altered protein levels diagnostic accuracy of new health technologies to decline over time as evidence from independent 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 biomarker 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

Table 2. Indicators for the assessment of methodological quality

<table>
<thead>
<tr>
<th>Domain</th>
<th>Patient selection</th>
<th>Index test</th>
<th>Reference standard</th>
<th>Flow and timing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description.</td>
<td>Describe methods of patient selection. Describe included patients (characteristics, prior testing, presentation, intended use of index test and setting).</td>
<td>Describe the index test (s) 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</td>
<td>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</td>
<td>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</td>
</tr>
</tbody>
</table>
### Table 2. Indicators for the assessment of methodological quality (Continued)

| 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, **No** if non-consecutive patients or a non-random sample of individuals were recruited, **Unclear** if patient selection was not clearly described. | Was calibration of examiners undertaken and results reported? Classify as **Yes** if the examiners did not participate in dedicated training and calibration was reported to an acceptable standard, **No** if the examiners did not participate in dedicated training or was not assessed, or training was undertaken but calibration was not to an acceptable standard, **Unclear** if the information on training and calibration was not stated. | 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 result is independently confirmed by at least two specialists, **No** if the test result is not independently confirmed by at least two specialists or there was lack of agreement between specialists, **Unclear** if the study does not state who confirmed the results. | 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, **No** if the delay between the index test(s) and reference standard is considered unacceptable for the majority of participants, **Unclear** if the delay between the index test(s) and reference standard is not explicitly stated. | Did the study avoid inappropriate exclusions? Classify as **Yes** if the sample consisted of apparently healthy individuals, **No** if only individuals with existing PMDs were recruited. **Unclear** if exclusions were not clearly stated. | 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, **No** if interpreters of index test results do not make any attempt to remain blinded. | Were the reference standard results interpreted without knowledge of the results of the index test? Classify as **Yes** if specialists clearly do not know index test results when performing the examination and clinical evaluation or evaluating follow-up. **No** if the same reference standard was used in all participants, **Unclear** if the same reference standard was not used in all participants. | Did all patients receive the same reference standard? Classify as **Yes** if the same reference standard was used in all participants, **No** if the same reference standard was not used in all participants, **Unclear** if it initial screening and reference standard (examination and clinical evaluation) is greater than 6 weeks then this will be considered an unacceptable delay |
Table 2. Indicators for the assessment of methodological quality  

<table>
<thead>
<tr>
<th>Description</th>
<th>Yes</th>
<th>No</th>
<th>Unclear</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpreters of index test results clearly know results of reference standard</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classify as <strong>Unclear</strong> if study did not provide any information on whether interpreters of index tests were blinded to reference standard</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-up data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classify as <strong>Yes</strong> if specialists clearly know index test results when performing the examination and clinical evaluation or evaluating follow-up data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classify as <strong>Unclear</strong> if study did not provide any information on whether specialists were blinded to the index test results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>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?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classify as <strong>Yes</strong> if index test results were interpreted without knowledge.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classify as <strong>No</strong> if the index test results were interpreted with knowledge.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classify as <strong>Unclear</strong> if it is unclear whether the results of the second index test were interpreted without knowledge of the results of the first index test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were all patients included in the analysis?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classify as <strong>Yes</strong> if all patients were included in the analysis.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classify as <strong>No</strong> if only some patients were included in the analysis.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classify as <strong>Unclear</strong> if it is unclear whether all patients were included in the analysis.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Risk of bias: High/Low/Unclear.</th>
<th>Could the selection of individuals have introduced bias?</th>
<th>Could the conduct or interpretation of the index test have introduced bias?</th>
<th>Could the reference standard, its conduct, or its interpretation have introduced bias?</th>
<th>Could the patient flow have introduced bias?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concerns regarding applicability: High/Low/Unclear.</td>
<td>Are there concerns that the included individuals do not match the review question?</td>
<td>Are there concerns that the index test, its conduct, or interpretation differ from the review question?</td>
<td>Are there concerns that the target condition as defined by the reference standard does not match the review question?</td>
<td></td>
</tr>
</tbody>
</table>

PMDs = potentially malignant disorders.

**APPENDICES**

**Appendix 1. MEDLINE via OVID search strategy**

1. exp Mouth/
2. Check/
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
Clinical assessment to screen for the detection of oral cavity cancer and potentially malignant disorders in apparently healthy adults
(Protocol)
Copyright © 2012 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.
HISTORY


CONTRIBUTIONS OF AUTHORS

Tanya Walsh (TW) and Joseph Liu (JL) wrote this protocol with contributions from Paul Brocklehurst (PB), Graham Ogden (GO), Crispian Scully (CS), Ross Kerr (RK), Mark Lingen (ML) and Saman Warnakulasuriya (SW).

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 Orthodontic Society, UK; British Society of Paediatric Dentistry, UK; National Center for Dental Hygiene Research & Practice, USA and New York University College of Dentistry, USA) providing funding for the editorial process (www.ohg.cochrane.org).
  - National Institute for Health Research (NIHR), UK.
  All reviews in the Cochrane Oral Health Group are supported by NIHR Systematic Reviews Programme infrastructure funding.