



# HHS Public Access

Author manuscript

*J Evid Based Dent Pract.* Author manuscript; available in PMC 2015 October 26.

Published in final edited form as:

*J Evid Based Dent Pract.* 2013 December ; 13(4): 151–154. doi:10.1016/j.jebdp.2013.10.006.

## Clinical Oral Examinations May Not Be Predictive of Dysplasia or Oral Squamous Cell Carcinoma

Jennifer L. Cleveland, DDS, MPH [Dental Officer/Epidemiologist] and Valerie A. Robison, DDS, MPH, PhD [Dental Officer Team Lead Valerie]

Surveillance, Investigation, and Research Team, Division of Oral Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, USA, Tel.:+1 770 488 6066

Jennifer L. Cleveland: JLCleveland@cdc.gov; Valerie A. Robison: VRobison@cdc.gov

### Summary

#### Selection Criteria

The investigators conducted a systematic search of PubMed, Web of Knowledge, and the Cochrane library from January 1, 1966 through January 20, 2010 using the search terms “oral mucosal lesion screening” and “oral lesions.” Additional articles were identified from other sources, such as reference lists and journals' Web sites. Data from screening and observational studies and randomized controlled trials published in English were considered. Studies were selected if they met the following criteria: (1) included histologic evaluation from biopsied lesions that were clinically detected, including some studies that used both clinical oral examinations and adjunctive techniques; (2) involved patients who sought care at either primary care medical or dental practices, were referred to a clinic because they had an oral mucosal disease, or received cancer therapy at a cancer treatment center; and (3) included patients who had either primary oral mucosal lesions or recurrent second oral malignancies not limited by stage or grade.

#### Key Study Factor

The authors conducted a systematic review and meta-analysis of studies assessing the effectiveness of clinical oral examinations (COEs) in predicting oral dysplasia or oral squamous cell carcinoma (OSCC). Quality of the studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool, which is an evidenced-based quality assessment tool used in systematic reviews of diagnostic accuracy studies.<sup>1</sup> QUADAS consists of 14 questions or criteria to which the possible responses are “yes,” “no,” or “unknown.” This tool was used to evaluate the quality of the studies, using criteria

**Purpose/Question:** To assess the effectiveness of the clinical oral exam in predicting potentially malignant epithelial lesions or oral squamous cell carcinomas

**Type of Study/Design:** Systematic review with meta-analysis of data

**Level Of Evidence:** Level 2: Limited-quality patient-oriented evidence

**Strength Of Recommendation Grade:** Grade B: Inconsistent or limited-quality patient-oriented evidence

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

such as representativeness of the study samples, eligibility criteria, study withdrawals, and whether patients received index testing (clinical oral examinations) and reference testing (gold standard test [biopsy]). QUADAS does not create an overall quality score but can be used to distinguish between high- and low-quality studies. The authors also used five of the QUADAS criteria to assess the level of the risk of bias (high, medium, and low).

### Main Outcome Measure

The primary outcome measure was a histologic confirmation of dysplasia or OSCC in an oral mucosal lesion submitted for biopsy. For each study, investigators reported that they calculated the sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR), and other measures of accuracy. The authors stated that because “clinically normal mucosa would not have been biopsied, 0.5 was added to all cells of the data analysis table to calculate the specificity.” (DOR is the odds of disease in test positives relative to the odds of disease in test negatives).<sup>2</sup> PLRs and NLRs state how many times more likely a patient is to have or not to have a disease given a particular test result.<sup>3</sup> PLRs above 5.0 and NLRs below 0.2 give strong diagnostic evidence,<sup>4</sup> while a value of 1.0 indicates that the diagnostic test provides no information on the probability of disease.

An overall meta-analysis was conducted for studies that met the inclusion criteria. Pooled summary measures for all studies combined were calculated for each statistical parameter. A random-effects model was used for the meta-analysis to account for inter-study variability. Heterogeneity between studies was assessed using the Cochran Q test and the inconsistency index,  $I^2$ , which describes the percentage of total variation across studies that is due to heterogeneity rather than chance.<sup>5</sup> (Heterogeneity is a measure of the inconsistency or variability between studies and can impact the generalizability of study results.) The authors also assessed publication bias.

### Main Results

Twenty-four observational studies met the inclusion criteria. The studies, which were published between 1997 and 2011, involved 7079 patients with an overall sample of 1956 patients who had lesions biopsied and histologically examined. Using QUADAS, the authors estimated that for two-thirds of the studies the risk of bias was low, and for one third, the risk was medium. No studies were determined to be at high risk for bias. The sensitivity, 0.93 (95% CI: 0.91-0.94), of the COE was high, indicating that the COE identified dysplasia or OSCC when disease was truly present (small number of false negatives). The specificity, 0.31 (95% CI: 0.28-0.34), was low, indicating that COE was poor in ruling out disease when disease was not present (large number of false positives). In this study, PLRs and NLRs were poor, indicating a relatively low ability of the COE to estimate the likelihood of disease in an individual. Similarly, the overall DOR, which can range from zero to infinity, was low (6.1; 95% CI: 2.1-17.60), suggesting that the COE was ineffective as a diagnostic method in predicting oral dysplasia or OSCC. Heterogeneity between studies was high, as indicated by the high  $I^2$  scores and corresponding  $P$ -values (all = 0.01), likely reflecting differences in patient populations and selection criteria. No publication bias was found.

## Conclusions

The authors used sensitivity, specificity, and other parameters to evaluate the COE as a diagnostic test. They concluded that diagnosis of oral dysplasia or OSCC based on a COE alone correlates poorly with the histological results from a biopsy. This discrepancy may be explained by the similarity in appearance of benign and dysplastic oral mucosal lesions, expertise of examiners, and variations in diagnoses of pathologists. The authors voiced concern about the potential for false negatives, in which disease is present but not detected and treated. They also voiced concern about the high number of false positives, which may lead to unnecessary worry for those told they might have pathology. The authors supported the development of more sensitive and specific adjunctive techniques to improve the accuracy of clinical detection and diagnosis of oral dysplasia and OSCC.<sup>6,7</sup>

## Commentary and Analysis

Systematic reviews of diagnostic and screening tests are fundamentally similar to other types of systematic reviews, but they differ in the criteria used to assess the quality of studies, the potential for bias, and the statistical methods used to combine results.<sup>3,8</sup> The purpose of this meta-analysis was to evaluate the effectiveness of the COE in predicting the histologic diagnosis (biopsy) of dysplasia or OSCC. However, while the COE can identify potentially malignant lesions, it also can identify other abnormalities that may need a biopsy and treatment.<sup>9</sup> The authors often referred to the COE as a diagnostic test, when in fact it is a screening test. Screening tests are intended to cast a broad net in capturing almost everyone in an asymptomatic population who has the disease.<sup>8</sup> Although some areas of the mouth are difficult to visualize (e.g., ventral surface of the tongue and floor of the mouth), one would expect a high sensitivity (i.e., few false negatives) for areas that are clearly visible. Therefore a useful screening test is one that can effectively rule out disease among those who test negative.<sup>10</sup> However, because oral cancer is such a rare disease,<sup>11</sup> it is reasonable to expect a high number of false positive results compared to diseases that are more common. In contrast, a diagnostic test determines whether disease is actually present.<sup>3</sup> In this study, the diagnostic test was the gold standard, biopsy, which should be adept at ruling in or detecting disease and should have a high specificity (i.e., few false positives).

In this analysis, the authors, however, included only persons with test-positive COEs involving detected or suspected lesions that had been referred for biopsy ( $n = 7079$ ) and who actually had the biopsy conducted ( $n = 1956$ ). These lesions were subsequently determined by histology to have or not to have true disease; however, the authors do not provide these numbers. Persons identified by COE to have clinically normal mucosa (test negatives) would not have been referred for a biopsy and were not included in the study sample. Because these patients were not followed to determine who remained disease-free or developed disease, we have no information about true negatives or false negatives. In addition, we have no information about persons who tested positive on COE, but did not proceed to the next step or were found on subsequent examination to have a lesion that did not require a biopsy (false positives). To account, in part, for the lack of information, the authors indicate that they added 0.5 in the cells for true negatives and false positives in order to calculate specificity. However, the authors did not specify the statistical methodology that

allows for this type of substitution. The pooled summary estimates for the indicators assessing the accuracy of the COE in identifying oral dysplasia or OSCC, therefore, could not be directly measured in this meta-analysis, so reported values for sensitivity, specificity, PLR, NLR, and DOR for this meta-analysis may not be valid.

The transparency of the analysis would be improved with the addition of relevant findings from each included study. These findings would include the sample size along with estimates (or assumptions in the absence of data) used to determine the number (percent) that were true positive (TP), false negative (FN), true negative (TN), and false positive (FP) and selected indicators. For example, it's unclear how the authors estimated sensitivity when they have no information on the percent of persons with a false negative test. In addition, while it could have been possible to examine the percent of persons with a positive screening test (i.e., COE) who were found on biopsy to have dysplasia or malignancy (i.e., positive predictive value), the total number of patients with a positive COE, not just those who had a biopsy, would be needed.

The degree of heterogeneity across the studies was significant, as indicated by the  $p$ -values of the  $I^2$  scores ( $p = 0.01$ ). In such cases, the degree of heterogeneity may be so large that it may be inappropriate to pool the performance of individual study parameters. When such heterogeneity across studies exists, it is difficult to draw conclusions about the summary test of specificity and the generalizability of results.

The authors used the QUADUS tool to evaluate the quality of the studies. The QUADUS tool includes 14 criteria to assess study quality, and the authors used five of the criteria to assign a level for the potential risk for bias (low, medium, or high) to each study.<sup>1</sup> In reviewing the results of the QUADUS tool for the 24 studies, it is the opinion of these reviewers that the level of bias and heterogeneity among the studies appeared higher than was reported by the authors. For example, for the first QUADUS criteria, "Was the spectrum of patients representative of the patients who will receive the test in practice?" the answer was "no" for 21 of 24 studies (87.5%). For a third of the studies, the selection criteria were not clearly described. Execution of the reference test (biopsy) was not described in detail to permit replication in 29.2% of studies. More importantly, for four of the studies, neither the whole sample nor a random selection of the sample appeared to have received the reference or gold standard biopsy, which the authors indicated was a condition for exclusion. In addition, the authors of this analysis did not describe how they used five of the QUADUS criteria to determine that a study was at low (70.8%), medium (29.2%), or high risk (0.0%) for bias. We suggest that future systematic reviews evaluating the effectiveness of methods to predict oral dysplasia or OSCC use the revised QUADUS-2 tool, which allows for more transparent rating of bias and applicability of primary diagnostic accuracy studies.<sup>12</sup>

Oral cancer screening is defined by the American Dental Association (ADA) "as the process by which an asymptomatic patient is evaluated to determine if he/she is 'likely' or 'unlikely' to have a potentially malignant or malignant lesion."<sup>9</sup> In a dental setting, the act of "screening" occurs when a patient reports for care and the practitioner obtains that patient's health history to assess risk, followed by the performance of a visual and tactile examination ... to detect *any* oral abnormality."<sup>9</sup> However, for a COE to be considered an effective

screening test for oral cancer, early diagnosis and treatment would have to show an impact on the course of the disease. At this point, the evidence is sparse regarding the effectiveness of COEs (or more accurately, oral cancer screening) in detecting potentially malignant or oral malignant lesions or for improving morbidity or mortality.<sup>9,13</sup> A nine-year randomized controlled trial provided some evidence that visual examination may have helped reduce death rates in certain high-risk patients, such as tobacco and heavy alcohol users.<sup>14</sup> Based on this evidence, the American Dental Association in 2010 encouraged practitioners to “remain alert for signs of potentially malignant lesions or early stage cancers while performing routine visual and tactile examinations,” particularly among tobacco and alcohol users.<sup>9</sup> Due to biases in the study, however, the U.S. Clinical Preventive Services Task Force concluded in 2004 and again in a recent draft Recommendation Statement that the evidence was insufficient to recommend for or against primary care providers routinely screening adults for oral cancer<sup>15</sup> and a recent Cochrane review found no evidence that visual oral examinations reduced death rates.<sup>16</sup>

In addition to the COE, the majority of studies in this systematic review used an adjunctive method, such as to-luidine blue and autofluorescence, in predicting dysplasia or OSCC among patients screening positive for suspected lesions. The authors suggest that the use of adjunctive applications to highlight such lesions may increase the accuracy of clinical diagnosis and diagnostic yield. However, the ADA, in their review of the evidence found that there is insufficient evidence that the commercially available devices based on tissue reflectance and auto fluorescence improve the detection of potentially malignant lesions beyond that of a conventional visual and tactile examination.<sup>9</sup>

The most definitive measure of the effectiveness of COE as a screening test would be a comparison of cause specific mortality rates among asymptomatic persons whose disease was picked up by the COE and those whose diagnosis was related to the development of symptoms. Because of factors related to costs, ethics and feasibility, such studies are not likely to be carried out. Current research shows promising developments in the use of biomarkers, such as salivary proteins and messenger RNAs, to discriminate patients with oral squamous cell carcinomas from healthy subjects.<sup>17</sup> However, until these technologies are validated in large clinical trials, the COE remains the primary screening test to detect potentially malignant and malignant lesions.

## Acknowledgments

The authors thank Dr. Barbara Gooch, Division of Oral Health, and Dr. Mona Saraiya, Division of Cancer Prevention, Centers for Disease Control and Prevention, Atlanta, for their comments and review of this article.

**Source of Funding:** None of the authors reported any external sources of funding to support this study

## References

1. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol.* 2003; 3:25. [PubMed: 14606960]
2. Glas AS, Lijmer JG, Prins MH, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol.* 2003; 56(11):1129–35. [PubMed: 14615004]

3. Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *Br Med J.* 2001; 323(7305):157–62. [PubMed: 11463691]
4. Eccles M, Freemantle N, Mason J. North of England evidence-based guideline development project: summary version of guidelines for the choice of antidepressants for depression in primary care. North of England Anti-depressant Guideline Development Group. *Fam Pract.* 1999; 16(2):103–11. [PubMed: 10381013]
5. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *Br Med J.* 2003; 327(7414):557–60. [PubMed: 12958120]
6. LeHew CW, Epstein JB, Kaste LM, Choi YK. Assessing oral cancer early detection: clarifying dentists' practices. *J Public Health Dent.* 2010; 70(2):93–100. [PubMed: 19765200]
7. LeHew CW, Epstein JB, Koerber A, Kaste LM. Training in the primary prevention and early detection of oral cancer: pilot study of its impact on clinicians' perceptions and intentions. *Ear Nose Throat J.* 2009; 88(1):748–53. [PubMed: 19172574]
8. Mallett S, Deeks JJ, Halligan S, et al. Systematic reviews of diagnostic tests in cancer: review of methods and reporting. *Br Med J.* 2006; 333(7565):413. [PubMed: 16849365]
9. Rethman MP, Carpenter W, Cohen EE, et al. Evidence-based clinical recommendations regarding screening for oral squamous cell carcinomas. *J Am Dent Assoc.* 2010; 141(5):509–20. [PubMed: 20436098]
10. Deeks J. Systematic reviews of evaluations of diagnostic and screening tests. *Br Med J.* 2001; 323:6.
11. Horner, MJ.; Ries, LAG.; Krapcho, M., et al. SEER Cancer Statistics Review, 1975-2006. National Cancer Institute; Bethesda, MD: 2009. [http://seer.cancer.gov/csr/1975\\_2006/](http://seer.cancer.gov/csr/1975_2006/), based on November 2008 SEER data submission, posted to the SEER web site
12. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011; 155(8):529–36. [PubMed: 22007046]
13. Downer MC, Moles DR, Palmer S, Speight PM. A systematic review of measures of effectiveness in screening for oral cancer and precancer. *Oral Oncol.* 2006; 42(6):551–60. [PubMed: 16815505]
14. Kujan O, Glenny AM, Duxbury J, Thakker N, Sloan P. Evaluation of screening strategies for improving oral cancer mortality: a Cochrane systematic review. *J Dent Educ.* 2005; 69(2):255–65. [PubMed: 15689610]
15. USPST Screening for Oral Cancer. 2004. <http://www.uspreventiveservicestaskforce.org/uspstf/uspso.html>
16. Brocklehurst P, Kujan O, Glenny AM, et al. Screening programmes for the early detection and prevention of oral cancer. *Cochrane Database Syst Rev.* 2010; (11):CD004150. [PubMed: 21069680]
17. Elashoff D, Zhou H, Reiss J, et al. Prevalidation of salivary biomarkers for oral cancer detection. *Cancer Epidemiol Biomarkers Prev.* 2012; 21(4):664–72. [PubMed: 22301830]