

IN BRIEF

- In order to gain an idea of the kind of results that screening and pilot programmes for oral cancer and precancer are producing, a review of these studies and a meta-analysis of their findings was undertaken.
- The studies analysed were conducted in varied circumstances regarding screening strategy used, populations involved, prevalence of disease in those populations, personnel undertaking screening, and their experience and training.
- In order to derive average numerical values for screening performance, a meta-analysis of the probabilities of screeners correctly identifying those with and without relevant target lesions, not known to have been used previously in the field of oral health was adopted.
- Despite the heterogeneity among studies, the ability of the screeners to discriminate between those with relevant lesions and those without appeared to be good in most instances. However, among patients referred for secondary care the uptake of available services was shown to vary enormously.

Meta-analysis of measures of performance reported in oral cancer and precancer screening studies

D. R. Moles,¹ M. C. Downer² and P. M. Speight³

Objective To elicit a range of values for sensitivity, specificity and other measures of performance in screening for oral cancer and precancer.

Method A literature search which included three databases was conducted. Strict inclusion criteria were applied. Values for sensitivity (Sn) and specificity (Sp), from seven investigations, were expressed as a receiver operator characteristic (ROC) curve. Meta-analysis of the combined results was used to produce a summary operator characteristic (SROC) curve.

Results The pooled weighted value of Sn from the seven studies was 0.796. From the SROC, the corresponding value of Sp at this level of Sn was 0.977 (95% CI 0.941, 0.991). When Sp was held at 0.977, the corresponding value of Sn from the SROC was 0.796 (95% CI 0.594, 0.912).

Conclusions The reports selected for eventual inclusion revealed a high level of heterogeneity with respect to the location of investigations, prevalence of lesions, the personnel used and other factors. The meta-analysis indicated that overall the studies had high discriminatory ability. The estimates of Sn and Sp, and values obtained for other measures of screening performance, were considered suitable for input to a simulation model in assessing the likely cost-utility of a variety of screening scenarios in further planned research.

During the past decade, there has been mounting interest in the possibility of instituting screening programmes for oral cancer and precancer. Although a disease of relatively low incidence, oral cancer has high morbidity and mortality and appears to fulfil many of the criteria of a disease suitable for screening population groups at risk.¹⁻³

Nevertheless, it would be difficult to envisage the formal acceptance of population screening for oral cancer as a part of health policy without its likely costs and benefits, as well as its feasibility and suitability, having been evaluated. The best evidence of the efficacy of a clinical intervention is provided by a randomised controlled trial (RCT), this being the 'gold standard' for evaluations of effectiveness. However, the cost and logistical difficulties of organising and managing a trial of screening for such a relatively uncommon disease would be formidable. A feasible alternative in such circumstances is the use of simulation modelling. This technique synthesises and analyses data collected from multiple sources, including the literature, and is capable of generating valid cost-effectiveness data. Simulation modelling has the added advantage that by the means of sensitivity analysis, it enables a range of screening scenarios to be examined relatively cheaply and the optimum approach identified. However RCT would allow at the most only a very few programmes to be evaluated and compared simultaneously and would be extremely costly to mount.

The use of a simple model to simulate opportunistic screening of patients at risk of oral cancer in general dental practice and provide a tentative determination of the health gain screening might achieve, together with its cost-effectiveness, has been reported previously.¹⁻³ However, the information on the validity of the clinical screening method used in these studies was from very limited sources and the economic data in particular were crude. Therefore, as a preliminary step towards collecting a comprehensive data set on all aspects of oral cancer and precancer screening with a view to constructing a more refined model capable of simulating and evaluating a variety of screening scenarios, it was decided to conduct a comprehensive literature search to assemble a range of reported values for various aspects of screening performance. These would include sensitivity, specificity and other expressions of the validity of clinical oral examination as a screening test, the uptake of screening programmes among target populations and compliance in attendance for follow up diagnostic examination in screen-detected cases. It was also determined that where appropriate, meta-analysis of the data would be undertaken.

It was decided that the collection of data on the cost of screening and follow-up diagnostic work, and on the cost of treatment and rehabilitation of oral cancer patients, should be the subject of a separate investigation.

¹Clinical Lecturer, Department of Oral Pathology, University College London, ²Emeritus Professor, Biostatistics Unit, Eastman Dental Institute, University College London, and Oral Health and Development, University Dental Hospital, Manchester, ³Professor, Department of Oral Pathology, University College London

Correspondence to: Dr D. R. Moles, Department of Oral Pathology, Eastman Dental Institute for Oral Health Care Sciences, University College London, 256 Gray's Inn Road, London WC1X 8LD, UK.

E-mail: d.moles@eastman.ucl.ac.uk

Refereed paper

Received 16.05.01; Accepted 10.10.01

© British Dental Journal 2002; 192: 340-344

Table 1 Criteria for inclusion and exclusion of studies

Types of study included	
1)	Prospective field studies concerned, <i>inter alia</i> , with the validation of clinical screening, or the performance of clinical examination-based screening programmes, for oral cancer and/or precancer.
2)	Original work presenting estimates of parameters relating to oral cancer and/or precancer screening performance, ie sensitivity, specificity, positive and/or negative predictive value, lesion prevalence in target group; response to screening invitation, compliance in follow up.
3)	Reviews providing secondary source data (for reference purposes only)
Types of study excluded	
4)	Investigations concerned with genetic markers or diagnostic aids (eg toluidine blue)
5)	Epidemiological surveys of oral cancer and/or precancer.

METHOD

The review was designed to be of low recall and high precision and the remit was limited to a consideration of studies that would yield quantitative information on specific parameters of oral cancer and precancer screening performance. Although not a systematic review as such, it was conducted as far as practicable in accordance with accepted guidelines for such studies.

Inclusion and exclusion criteria

In order to keep the review within manageable bounds and focused on the essential information sought, only research reports within the area of oral cancer and precancer were admitted. Moreover these had to relate specifically to screening, or to screening programmes, and assess the validity of the screening activity using at least a 'soft' gold standard. Ideally this would comprise a comprehensive diagnostic work up of all subjects screened positive, but failing that, detailed clinical examination by an oral medicine or surgery specialist would suffice. In order to provide estimates of true- and false-negative rates it was determined that studies accorded full weight should also include the same diagnostic information on at least a proportion of those screened negative. However, studies from which it was possible to derive only true- and false-positive rates, among those screened and referred to secondary care facilities for further investigation, were also deemed eligible for inclusion. These were collated and examined separately. The inclusion and exclusion criteria are summarised in Table 1.

Added value would be gained where study reports provided useful additional information. This would include the type of personnel involved in screening (and details of their training and examiner reproducibility where applicable); a full list of the target lesions classified as positive and those regarded as negative; effect modifiers if specified such as age, gender, tobacco usage and drinking habits of the target group; and whether programmes were invitational or opportunistic. The inclusion of multiple studies based on the same data was avoided. A full list of studies both included in, and excluded from the review was maintained.

Search strategy

A literature search was conducted using the databases PubMed, SCISEARCH and the Cochrane Library. Subject headings specified were 'oral neoplasms', 'oral cancer', 'oral precancer' 'screening' and 'screening programmes'. This yielded a collection of 60 articles of *prima facie* interest in peer-reviewed journals published up until the end of December 2000. Abstracts displayed indicated that 27 of these papers might be worthy of detailed scrutiny. A further three papers of possible interest, not cited in the databases, were disclosed from hand searching of journals, or through personal contact with other investigators working in the field.

Data synthesis

Global estimates for sensitivity and specificity were obtained from the selected studies using the summary receiver operator characteristic (SROC) curve meta-analytical technique described by Irwig *et al.*⁴ The first step of this technique was to utilise the odds ratios from the studies as a measure of the discriminatory ability of the diagnostic procedures. These odds ratios were combined using a standard random-effects meta-analysis within the STATA software package.⁵ The meta-analysis results were checked for the possibility of skew according to the methods described by Moses and co-workers by use of a weighted least squares regression analysis.^{6,7} A SROC curve was constructed from the pooled odds ratio (with 95% confidence intervals) by calculating the values of specificity for every possible value of sensitivity. Finally a weighted 'pooled' value for sensitivity was obtained from the studies and the corresponding specificity value (with 95% confidence intervals), for this level of sensitivity, was calculated from the equation of the SROC curve. This value of specificity was utilised to calculate the corresponding point estimate and 95% confidence interval for sensitivity.

RESULTS

Six studies captured in the literature search, describing seven substantive or pilot screening programmes, yielded data on the sensitivity and specificity of the specified screening procedure (systematic visual examination of the oral mucosa) or provided sufficient information for these values to be derived.⁸⁻¹⁴ The studies and the basic information they yielded are listed in Table 2. Two of these, and the further reports listed in Table 3, provided outcome data on patients referred for secondary care including numbers of individuals screened, numbers referred, numbers who attended, true- and false-positive rates, and positive predictive values of the test.^{9,12,15-21}

With regard to the findings of the first group of studies (Table 2), sensitivity (Sn) values ranged from 0.60 to 0.95. Specificity (Sp) values were at least 0.94 apart from the Sri Lanka study where the basic health workers returned a false-positive rate of 19% (Sp = 0.81).¹³ In the second group of studies (Table 3), positive predictive values (PPV) among screened individuals attending secondary care facilities, fluctuated widely between 0.45^{12,15} and 1.00 in the study of Field *et al*, where prevalence was evidently extremely low and only four cases with precancerous lesions were referred and subsequently attended.¹⁶

Table 2 Summary of values for expressions of the validity of screening for oral cancer and precancer derived from seven studies

Study and location	Downer <i>et al.</i> (1995) London England	Ikeda <i>et al.</i> (1995) Nagoya Japan	Jullien <i>et al.</i> (1995) (GP) London England	Jullien <i>et al.</i> (1995) (Hosp) London England	Mathew <i>et al.</i> (1997) Kerala India	Mehta <i>et al.</i> (1986) Kerala India	Warakulasuriya and Pindborg (1990) Sri Lanka
Screened n	309	154	985	1042	2069	1921	1872
Prevalence %	5.50	9.74	2.23	3.07	0.25	1.41	21.63
Sensitivity	0.71	0.60	0.64	0.81	0.94	0.59	0.95
Specificity	0.99	0.94	0.99	0.99	0.98	0.98	0.81
PPV	0.86	0.50	0.64	0.68	0.87	0.31	0.58
NPV	0.98	0.96	0.99	0.99	0.99	0.99	0.98

Table 3 Summary of data from nine studies relating to follow up diagnosis of screened patients referred to secondary care facilities

Study & location	Screened n	Referred n	Attended n %	True + n	False + n	PPV
Fernandez Garrote <i>et al.</i> (1995) Cuba	13 x 10 ⁶	30,244	8,703 28.8	3,924	4,779	0.45
Field <i>et al.</i> (1995) England	1,947	4	4 100	4	0	1.00
Ikeda <i>et al.</i> (1995) Japan	802	32	25 78.1	20	5	0.80
Ikeda <i>et al.</i> (1991) Japan	3,131	77	37 48.1	27	10	0.73
Mehta <i>et al.</i> (1986) India	39,331	523	377 72.1	169	208	0.45
Nagao <i>et al.</i> (2000) Japan	19,056	200	137 68.5	73	21	0.78
Sankaranarayanan <i>et al.</i> (2000) India	49,179	3,585	1,877 52.4	1,346	531	0.72
Warnakulasuriya & Nanayakkara (1991) Sri Lanka	57,124	3,559	2,193 61.6	1,741	431	0.80
Warnakulasuriya <i>et al.</i> (1988) Sri Lanka	29,295	1,230	664 54.0	412	252	0.62

Table 4 Random effects meta-analysis of screening test discriminatory ability of studies listed in Table 2

Study	Weight	Log Odds Ratio	95% CI	
			Lower	Upper
Ikeda	0.53	3.08	1.84	4.31
Mehta	0.60	4.35	3.51	5.18
Warnakulasuriya	0.65	4.37	3.91	4.83
Jullien (GP)	0.56	5.34	4.23	6.45
Downer	0.44	5.85	4.11	7.59
Jullien (Hosp)	0.57	5.89	4.83	6.94
Mathew	0.63	6.89	6.21	7.57
Combined		5.11	4.12	6.09

Meta-analysis

Figure 1 shows the ROC curve for the seven studies in Table 2 which were included in the meta-analysis. Figure 2 and Table 4 show the results of the random effects meta-analysis of the discriminatory ability of the screening studies. The test of skew gave a coefficient for the effect of skew on discrimination of 1.64 (95% CI -4.01, 7.30) with a P-value of 0.49, indicating no evidence to suggest that the SROC curve was asymmetrical. The SROC curve for these studies is shown in Figure 3. The weighted pooled value of their sensitivity was 0.796. From the equation for the SROC curve, the corresponding value of specificity at this level of sensitivity was 0.977 (95% CI 0.941, 0.991). When specificity was held at 0.977, the corresponding value of sensitivity from the SROC curve was 0.796 (95% CI 0.594, 0.912).

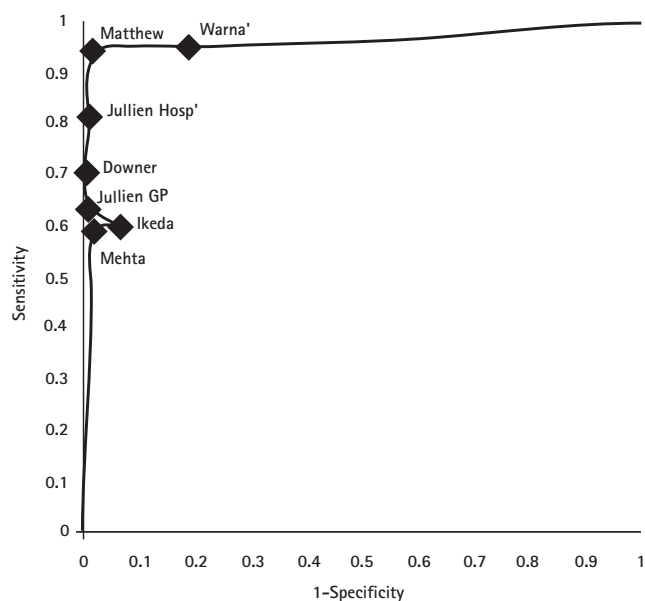


Figure 1. ROC curve for seven oral cancer screening studies

DISCUSSION

The studies listed in Tables 2 and 3 which met the inclusion criteria (Table 1) showed considerable heterogeneity with respect to the country and specific location where they were conducted; the type of personnel carrying out the screening and their training and calibration; the demographic characteristics of the target population; the numbers screened; the types of lesions categorised as 'positive' (target lesions); the prevalence of target lesions; the referral centres and type of personnel representing the gold standard examiners; and whether the programmes were opportunistic or invitational. For example, with regard to the British studies in the first group (Table 1), that of Downer *et al*⁸ conducted in a company headquarters, and the general practice (GP) component of the study of Jullien *et al.*¹⁴ were invitational, as was the study of Ikeda *et al.* in part of the Nagoya conurbation in Japan.⁹ The hospital component of the study of Jullien *et al.*, however, was opportunistic.¹⁰ Mathew *et al* provided an interim report of a controlled trial of screening, randomised by administrative district, in Kerala, India where there is a high oral cancer incidence.¹¹ Mehta *et al* described a house-to-house case finding exercise, also in Kerala.¹² The reports of Warnakulasuriya and colleagues in Sri Lanka described a similar approach.^{13,20,21} The British and Japanese studies employed dentists from various practice settings as screeners, without specific training in the first instance and with training and calibration in the second. On the other hand, the studies in the Indian sub-continent, reflecting the economic circumstances prevailing in that part of the world, investigated the use of specifically trained basic health workers as screeners in a possible relatively low cost preventive strategy for combating their exceptionally high levels of the disease. In practical terms, the clear heterogeneity among the studies appeared to have only a moderate influence on the sensitivity and specificity values reported. The discriminatory ability demonstrated by the screening personnel, irrespective of their grade and training, was generally of a fairly high order and did not appear to be greatly affected by the varied circumstances of the studies.

It is rare that the results of screening tests are combined using formal methods such as meta-analysis. It is even less common to see such analyses reported within the dental literature. There are a variety of inappropriate methods that have been employed in attempts to synthesise the results from multiple screening studies. It therefore seems worthwhile to discuss some of the methodological issues that are relevant to the synthesis of data relating to diagnostic tests.

Studies evaluating screening programmes may differ in their thresholds for calling a test result positive, in the present instance possibly as a result of variation in the target lesions specified or systematic variation resulting from training of the screeners, or lack of it. Since the sensitivity and specificity of a screening test are inter-related, changes to the threshold for a positive diagnosis will affect both measures. Thus if investigators consider it essential to detect as many positive cases as possible, the sensitivity of a screening test could theoretically be improved, but only at the cost of reducing

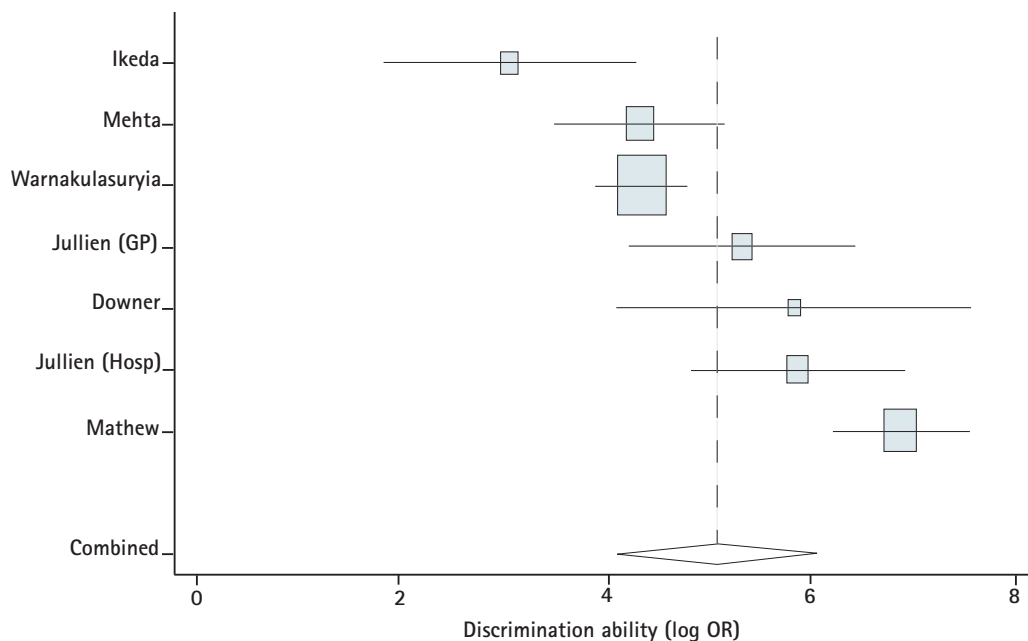


Figure 2 Random effects meta-analysis of discriminatory ability of seven oral cancer screening studies

specificity. This variation in positive threshold can be usefully summarised with a receiver operator characteristic (ROC) curve. This is achieved by plotting the sensitivity of a test against 1-specificity.

A method for combining the results of several studies must account for both the discriminatory ability of each study and the variation in diagnostic threshold. It is inappropriate to directly pool the results of each investigation since the differing prevalence of positive lesions across studies acts as a confounding factor when diagnostic thresholds differ. This is particularly likely to be a problem if there is a wide range of prevalence of disease across different studies (eg if trying to combine studies from different populations such as the United Kingdom and India). It is also inappropriate to calculate sensitivity and specificity separately within each study and to attempt to derive a weighted average of each measure (eg with weights being based on study size). This avoids the problem of confounding, but may still lead to an underestimate of the true accuracy if there is variation in the diagnostic thresholds used by different investigators.⁴

Since sensitivity and specificity are inter-related and since both diagnostic thresholds and disease prevalence may vary, the correct approach to combining the results of several studies is to use meta-analysis to produce a summary receiver operator characteristic curve (SROC) as previously described. If the odds ratio is constant across different thresholds this will lead to a symmetric SROC curve. Thus, if positive and negative screening results change the odds equally, then the diagnostic threshold is not skewed (biased) toward either diagnosis. To gain a global estimate of sensitivity and specificity it is necessary to derive a weighted estimate for one measure and then calculate the corresponding value (and confidence interval) for the other measure using the results of the meta-analysis.

The meta-analysis described is considered to have yielded an appropriate range of values, for the particular variables examined, for input to a simulation model. With the addition of other necessary data and the use of sensitivity analysis, this could then be used to ascertain the likely cost-utility of different oral cancer and precancer screening scenarios. This will be the subject of further research.

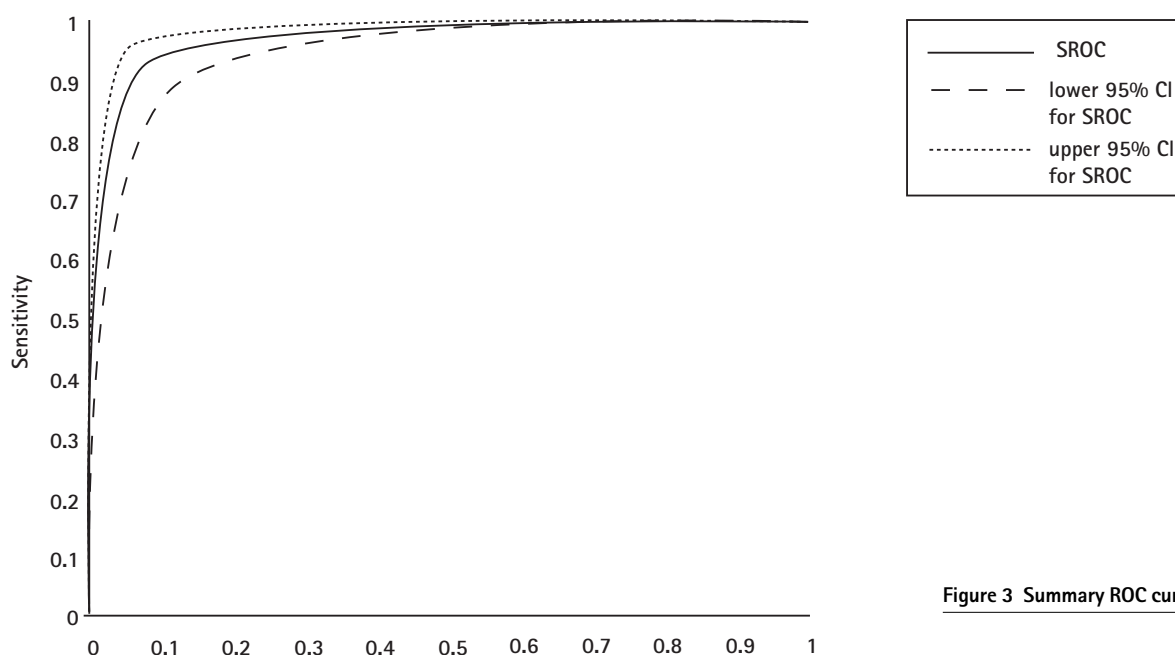


Figure 3 Summary ROC curve for seven studies

The authors would like to thank Mr Christopher Frost, Senior Lecturer in Medical Statistics, London School of Hygiene and Tropical Medicine for his assistance in obtaining details of an appropriate meta-analytical technique for screening tests.

1. Downer M C, Jullien J A, Speight P M. An interim determination of health gain from oral cancer and precancer screening: 1. Obtaining health state utilities. *Community Dent Health* 1997; **14**: 139-142.
2. Downer M C, Jullien J A, Speight P M. An interim determination of health gain from oral cancer and precancer screening: 2. Developing a model of population screening. *Community Dent Health* 1997; **14**: 227-232.
3. Downer M C, Jullien J A, Speight P M. An interim determination of health gain from oral cancer and precancer screening: 3. Preselecting high risk individuals. *Community Dent Health* 1998; **15**: 72-76.
4. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol* 1995; **48**: 119-130.
5. Stata Corporation. *STATA version 6*. Texas: Stata Corporation 1999.
6. Littenberg B, Moses L E. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. *Med Decis Making* 1993; **13**: 313-321.
7. Moses L E, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med* 1993; **12**: 1293-1316.
8. Downer M C, Evans A W, Hughes Hallett C M, Jullien J A, Speight P M, Zakrzewska J M. Evaluation of screening for oral cancer and precancer in a company headquarters. *Community Dent Oral Epidemiol* 1995; **23**: 84-88.
9. Ikeda N, Downe M C, 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 Dent Health* 1995; **12**: 133-137.
10. Jullien J A, Downer M C, Zakrzewska J M, Speight P M. Evaluation of a screening test for the early detection of oral cancer and precancer. *Community Dent Health* 1995; **12**: 3-7.
11. Mathew B, Sankaranarayanan R, Sunilkumar K B, Kuruvila B, Pisani P, Nair M K. Reproducibility and validity of oral visual inspection by trained health workers in the detection of oral cancer and precancer. *Br J Cancer* 1997; **76**: 390-394.
12. Mehta F S, Gupta P C, Bhonsle R B, Murti P R, Daftary D K, Pindborg J J. Detection of oral cancer using basic health workers in an area of high oral cancer incidence in India. *Cancer Detect Prev* 1986; **9**: 219-225.
13. Warnakulasuriya S, Pindborg J J. Reliability of oral precancer screening by primary health care workers in Sri Lanka. *Community Dent Health* 1990; **7**: 73-79.
14. Jullien J A, Zakrzewska J M, Downer M C, Speight P M. Attendance and compliance at an oral cancer screening programme in a general medical practice. *Oral Oncology Eur J Cancer* 1995; **31B**: 202-206.
15. Fernandez Garrote L, Sankaranarayanan R, Lence Anta J J, Rodriguez Salva A, Parkin D M. An evaluation of the oral screening program in Cuba. *Epidemiol* 1995; **6**: 428-431.
16. Field E A, Morrison T, Darling A E, Parr T A, Zakrzewska J M. Oral mucosal screening as an integral part of routine dental care. *Br Dent J* 1995; **179**: 262-266.
17. Ikeda N, Ishii T, Iida S, Kawai T. Epidemiological study of oral leukoplakia based on mass screening for oral mucosal diseases in a selected Japanese population. *Community Dent Oral Epidemiol* 1991; **19**: 160-163.
18. Nagao T, Ikeda N, Fukano H, Miyazaki H, Yano M, Warnakulasuriya S. Outcome following a population screening programme for oral cancer and precancer in Japan. *Oral Oncology* 2000; **36**: 340-346.
19. Sankaranarayanan R, Mathew B, Jacob B J, *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**: 664-673.
20. Warnakulasuriya K A A S, Nanayakkara B G. Reproducibility of an oral cancer and precancer detection program using a primary health care model in Sri Lanka. *Cancer Detect Prev* 1991; **15**: 331-334.
21. Warnakulasuriya S, Ekanayake A, Stjernsward J, Pindborg J J, Sivayoham S. Compliance following referral in the early detection of oral cancer and precancer in Sri Lanka. *Community Dent Oral Epidemiol* 1998; **16**: 326-329.

AMENDMENT

We would like to inform readers that an incorrect dosage was printed in a recent research paper. The dosage in the discussion of the paper in BDJ 2002; 192: 163, was wrongly given as 1:18000 and should have been 1:80000.