Association between oral leukoplakia and sex, age, and tobacco habits

B. ROED-PETERSEN,[1] P. C. GUPTA,[2] J. J. PINDBORG,[1,3] & B. SINGH [4]

The association between oral leukoplakia and possible etiological factors in patients at a Bombay hospital was analysed by a new multivariate computer analysis based on information theory. It was found that the most important etiological factors were bidi smoking and age. The well known sex–leukoplakia relationship was found to be secondary to sex differences in tobacco habits. The topographical distribution of leukoplakias within the oral cavity was also determined mainly by tobacco habits. It was concluded that the etiological factors analysed could not fully explain the presence or absence of leukoplakia but that other factors, so far unrecognized, may also play a role.

A preliminary report on the prevalence of leukoplakia in the present study sample was presented by Pindborg et al. (1965). This paper presents the results of a new statistical approach to the analysis of associations between oral leukoplakia and possible etiological factors, taking into consideration the ties that may exist between the independent variables.

POPULATION AND METHODS

The study population was made up of 9 982 patients arriving consecutively at the admission clinic of the Government Dental College in Bombay, India, between December 1963 and May 1964.

Leukoplakia was defined as any well demarcated, elevated white patch 5 mm or more in diameter that could not be scraped off and could not be attributed to the presence of other diseases. The definition carried no histological connotation.

Before being submitted to oral examination the patients were interviewed as to age, religion, and smoking and chewing habits, the composites of which have been described by Pindborg et al. (1967). The oral examination was carried out by electric

light, the oral cavity being screened by means of two mouth mirrors. Lesions complying with the above mentioned definition were registered together with details of their location.

Statistical analysis

The anamnestic and clinical findings were transferred to punch cards and the data were processed by computer.

Larsen (1968) developed an effective multivariate computer programme for the screening of discrete variables. It is well known that in cases where the outcome of any dependent variable is to be associated with a number of independent variables, any one of these may be expected to have no influence, to have an influence by itself, or to have an influence as a part of a multivariable association, or it may be expected that any association with the dependent variable is the mere result of secondary ties to one or more of the other independent variables. In the latter case the actual independent variable carries no information by itself and can be discarded, possibly thus unmasking hitherto unrecognized effects of other independent variables. If an analysis that takes into account the above mentioned possibilities of associations depends on the conventional nonparametric tests, such an analysis is practically impossible unless the amount of material is extremely large and/or the variables are evenly distributed within the material.

The analysis described by Larsen (1968) is based on the mathematical philosophy of cybernetics and in-

formation theory. Four terms are introduced among which the term more-information is new and is the key to the effectiveness of the screening method:

Entropy, $\hat{H}(A)$, is the variation of a variable A:

$$\hat{H}(A) = \sum_A \hat{p}(a)\ln\frac{1}{\hat{p}(a)} + C_1$$

where $\hat{p}(a)$ is the estimated probability for the occurrence of the single value 'a' of A, and C_1 is the correction for noncentrality of this estimate:

$$C_1 = \frac{k(A)-1}{2N}$$

where $k(A)$ is the number of single values in A, and N is the sample size.

Transinformation, $\hat{I}(A;B)$, is an estimate of the association between two variables A and B:

$$\hat{I}(A;B) = \sum_{AB} \hat{p}(a,b)\ln\frac{\hat{p}(a,b)}{\hat{p}(a)\,\hat{p}(b)} + C_2$$

where $\hat{p}(a,b)$ is the estimated probability for the occurrence of the combination of the single values of a and b.

$$C_2 = -\frac{(k(A)-1)(k(B)-1)}{2N}$$

which is proportional to the degrees of freedom:

$$df = (k(A)-1)(k(B)-1)$$

Conditioned transformation, $\hat{I}(A;B|C)$, is an estimate of the association between two variables A and B, which cannot be explained by a third variable C.

$$\hat{I}(A;C|B) = \sum_{ABC} \hat{p}(a,b,c)\ln\frac{\hat{p}(a,b|c)}{\hat{p}(a|c)\,\hat{p}(b|c)} + C_3$$

$$C_3 = -\frac{(k(A)-1)(k(B)-1)\,k(C)}{2N} = \frac{-df}{2N}$$

The entropy and transinformations are positive quantities. The estimates may, however, be slightly negative, because of the correction factors: $0 \leq \hat{I}(A;B) \leq \min(H(A);H(B))$.

Table 1. Distribution of data for 15 variables related to the etiology of oral leukoplakia

Variable	Classes			
	I	II	III	IV
leukoplakia (I = present, II = absent)	291	9 691		
sex (I = females, II = males)	3 295	6 687		
age (years) last birthday (I = 0–19, II = 20–39, III = 40–59, IV = 60–89)	2 104	4 462	2 798	618
religion (I = Hindu, II = Moslem, III = Christian, IV = other religions)	6 540	1 977	1 073	392
bidi smoking (I = +, II = —)	1 617	8 365		
cigarette smoking (I = +, II = —)	1 108	8 874		
hooka smoking (I = +, II = —)	36	9 946		
chilum smoking (I = +, II = —)	21	9 961		
cigar or pipe smoking (I = +, II = —)	16	9 966		
occasionally chewing pan without tobacco (I = +, II = —)	553	9 429		
chewing betel (I = +, II = —)	178	9 804		
chewing pan without tobacco (I = +, II = —)	323	9 659		
chewing pan with tobacco (I = +, II = —)	948	9 034		
chewing tobacco (I = +, II = —)	515	9 467		
bidi (I = bidi ± other habits, II = other habits excluding bidi, III = no habits)	1 617	2 825	5 540	

More-information, $\hat{I}(A,B,C)$, is an estimate of that part of the association between three variables A, B, and C that is not found in the three marginal distributions with two variables. The more-information may be negative because of the fact that some association between the three variables may be explained by other associations between them. A positive more-information shows that there is a so-called multivariate association between the three variables.

$$\hat{I}(A,B,C) = \sum_{ABC}\hat{p}(a,b,c)\ln\frac{\hat{p}(a,b,c)\hat{p}(a)\hat{p}(b)\hat{p}(c)}{\hat{p}(a,b)\hat{p}(a,c)\hat{p}(b,c)} + C_4$$

$$C_4 = -\frac{\{k(A)-1\}\{k(B)-1\}\{k(C)-1\}}{2N}$$

It can be shown that:

$$\hat{I}(A;B/C) = \hat{I}(A;B) + \hat{I}(A,B,C)$$

To test the significance of transinformation values a likelihood quotient test was carried out on the contingency tables resulting in a test value, Z, equal

Table 2. Entropy for the dependent variable leukoplakia and transinformations and probabilities for non-associations between the independent and the dependent variable

Independent variables	Dependent variable—leukoplakia: entropy 0.132	
	Trans-information	Probability
sex	0.007	<0.01 [a]
age	0.011	<0.01 [a]
religion	0.001	<0.01
bidi smoking	0.023	<0.01 [a]
cigarette smoking	0.000	0.22
hooka smoking	0.001	<0.01
chilum smoking	0.000	<0.01
cigar or pipe smoking	0.000	0.49
occasionally chewing pan without tobacco	0.001	<0.01
chewing betel	0.000	0.24
chewing pan without tobacco	0.000	0.15
chewing pan with tobacco	0.002	<0.01
chewing tobacco	0.000	0.01
bidi smoking (redefined)	0.028	<0.01 [a]

[a] Statistically significant associations between independent and dependent variables in cases where the transinformation value is greater than 0.005 and cannot be explained by more-information values to other independent variables.

to the transinformation estimate, without correcting factor, multiplied by twice the sample size. Z is distributed approximately according to the chi-square distribution.

In a similar way a test for the more-information was made. It seems likely, that this test is correct even though it has not yet been definitely proved.

Finally a chi-square test with Yates' correction was used for contingency tables that were not generated as part of the multivariate analysis.

In statistical testing, type 1 errors imply the finding of significant associations that have arisen by chance and that have no biological basis. In a study such as the present one in which large numbers of associations are tested the risk of type 1 errors becomes very important. To minimize this risk the 1 % level of significance was used.

RESULTS

Etiology

Leukoplakia was found in 291 (2.9%) of the 9 982 persons examined (Table 1). An evaluation of the transinformation values of the 13 independent variables of Table 2 revealed that 10 of these transinformations were negligible, being 0.000, 0.001, or 0.002. These 10 variables were therefore excluded from the information analysis even though 6 of them showed significant association with the dependent variable as determined from the corresponding contingency tables. Thus, on the basis of the numerical size of the transinformation values only bidi smoking, age, and sex should be considered as part of the information system related to oral leukoplakia. However, in order not to lose the small amounts of information contained in the other habit variables, bidi was redefined as shown at the bottom of Tables 1 and 2. This redefined variable for bidi smoking will be used in the following discussion.

Once the independent variable bidi smoking was given the more-information value related to leukoplakia and age was found to be $\hat{I}$(leukoplakia, bidi smoking, age) = -0.008 and to leukoplakia and sex $\hat{I}$(leukoplakia, bidi smoking, sex) = $-0,006$, whereas the more information value between leukoplakia, sex and age was only -0.002. Therefore:

$\hat{I}$(leukoplakia;age/bidi smoking)
= 0.011 $-$ 0.008 = 0.003

and,

$\hat{I}$(leukoplakia;sex/bidi smoking)
= 0.007 $-$ 0.006 = 0.001

according to the transformations in Table 2. Thus,

Table 3. Association between the independent variables bidi smoking and age and the dependent variable oral leukoplakia

Bidi smoking	Age (years)	Oral leukoplakia +	Oral leukoplakia —
bidi smoking ±other habits	0–19	0	17
	20–39	57 (8.49 %)	614
	40–59	124 (15.68 %)	667
	60–89	27 (19.57 %)	111
other habits excluding bidi	0–19	1 (0.71 %)	139
	20–39	26 (1.71 %)	1 492
	40–59	37 (3.73 %)	956
	60–89	9 (5.17 %)	165
no habits	0–19	0	1 947
	20–39	4 (0.18 %)	2 269
	40–59	3 (0.30 %)	1 011
	60–89	3 (0.98 %)	303

even though much of the age variation and nearly all of the sex variation of oral leukoplakia seemed to be explained once the bidi smoking habit was present in the information system, some unexplained variation still remained.

This is brought out in Table 3 where it is seen that within each of the three habit groups the proportion of oral leukoplakias increased systematically with age. In contrast, the age distribution differed significantly between the habit groups and the no-

Table 4. Association between the independent variables bidi smoking and sex and the dependent variable oral leukoplakia

Bidi smoking	Sex	Oral leukoplakia +	Oral leukoplakia —
bidi smoking ±other habits	females	7 (16.67 %)	35
	males	201 (12.76 %)	1 374
other habits excluding bidi	females	5 (0.86 %)	579
	males	68 (3.03 %)	2 173
no habits	females	4 (0.15 %)	2 665
	males	6 (0.21 %)	2 865

Table 5. Data distribution for four variables related to the topographical location of oral leukoplakia among 291 patients with lesions

Variable	Classes [a]						
	I	II	III	IV	V	VI	VII
Location	70	77	44	80	3	12	5
Habit	10	180	44	57			
Sex	16	275					
Age	1	87	164	39			

[a] Explanation of classes:

Location: I = commissures only;
II = buccal mucosa or lips only;
III = locations other than I and II only;
IV = combination of commissures and buccal mucosa or lips;
V = combination of commissures and locations other than I and II;
VI = Combination of buccal mucosa or lips and locations other than I and II;
VII = combinations of locations I, II and III.

Habit: I = no habit;
II = smoking habit only;
III = chewing habit only;
IV = combined smoking and chewing habit.

Sex: I = females;
II = males.

Age: I =   0–19 years;
II = 20–39 years;
III = 40–59 years;
IV = 60–89 years.

habit group (P< 0.01), the percentage of habitués in the age group 60–89 years being significantly lower (P< 0.01) than in the age group 40–59 years.

Similarly, as shown in Table 4, no sex variation in prevalence of leukoplakia could be demonstrated among bidi smokers (0.70 >P >0.50) or among non-smokers (0.90 >P >0.80), but among persons with single or mixed habits other than bidi, a sex variation in leukoplakia remained, which could not be explained by the other variables. No attempt was made to

Table 6. Entropy for the dependent variable location of oral leukoplakia and transinformations and probabilities for nonassociations between the independent and the dependent variables

Independent	Dependent variable—location: entropy 1.594	
	Transinformation	Probability
sex	0.025	0.20
age	0.020	0.01
habit	0.095	<0.01

Table 7. Location of 523 oral leukoplakias in relation to tobacco habits

Habit	Commissures		Buccal mucosa		Other locations		Total No.
	No.	% of group	No.	% of group	No.	% of group	
no habit	3	25.0	6	50.0	3	25.0	12
smoking habit	161	47.1	123	36.0	58	17.0	342
chewing habit	13	21.7	43	71.7	4	6.7	60
mixed chewing and smoking habit	47	43.1	47	43.1	15	13.8	109
total	224		219		80		523

break down further this habit variable because of the very small numbers in the cells of the contingency tables at this stage of the analysis.

Location of leukoplakias

On the basis of the findings in Tables 1 and 2, sex, age, and smoking and chewing habits were selected as independent variables for an analysis of factors determining the location of leukoplakias within the oral cavity. The results are presented in Tables 5 and 6. Again the transinformation value for the tobacco habit is far greater than those for sex and age. As more of the expected frequencies of the relevant contingency tables were less than 1, the probabilities given in Table 6 should be regarded with due reser-

Table 8. Location of lesions in 291 patients with oral leukoplakia

Location of lesion		Number	Percentage	
labial commissure	right	122	23.3	42.8
	left	102	19.5	
buccal mucosa	right	128	24.5	41.9
	left	91	17.4	
lip	upper	1	0.2	2.1
	lower	10	1.9	
alveolar ridge		35	6.7	
palate		18	3.4	
tongue		13	2.5	
gingiva		2	0.4	
floor		1	0.2	
total		523	100.0	

vation. The presence of positive more-information values for each combination of independent variables indicated a multivariable influence of these factors upon the dependent variable.

Among the 291 patients with oral leukoplakia a total of 523 leukoplakias were found. The proportions of patients with 1, 2, 3, 4, and 5 leukoplakias were 52.9%, 27.1%, 7.9%, 11.3%, and 0.7%, respectively. A cross-tabulation of location and smoking and chewing habits (Table 7) showed that the most frequent site in the smoking habit group was the commissures (47.1%) and in the chewing habit group the buccal mucosa (71.7%). The difference between the two habit groups was significant ($P < 0.01$). Among persons with both smoking and chewing habits the commissures and the buccal mucosa were evenly represented (43.1%).

Table 8 gives the exact distribution of the leukoplakias according to location. The labial commissures were the regions most often involved (42.8%) closely followed by the buccal mucosa (41.9%) It should be noticed that only 1 out of 11 cases of lip involvement was found in the upper lip, and that in 1 instance only was a lesion found in the floor of the mouth. In the two groups with leukoplakias on the commissures and buccal mucosa there seemed to be a tendency for more frequent involvement of the right side. This special point is more clearly seen in Table 9 which includes both unilateral and symmetrical bilateral cases. In the groups with unilateral involvement the difference between right and left side was significant for the buccal mucosa ($P < 0.01$), but not for the commissures ($P > 0.01$). Table 9 also demonstrates, that symmetrical bilateral involvement seemed to be more common for the commissures than for the buccal mucosa, but the difference was not statistically significant ($P > 0.01$).

Table 9. Distribution of 325 patients with leukoplakias of the commissures and buccal mucosa according to symmetry of lesions

Location	Single lesion right	Single lesion left	Symmetrical lesions	Total
commissures	56 (35.4 %)	36 (22.8 %)	66 (41.8 %)	158
buccal mucosa	76 (45.5 %)	39 (23.4 %)	52 (31.1 %)	167
total	132	75	118	325

DISCUSSION

Comparing the findings of the present study with those of earlier investigations of the etiology of oral leukoplakia, many of the differences may be ascribed to the new statistical approach applied in the present study.

The analysis based on information theory showed that in this study bidi smoking was the factor carrying the greatest information about the variation of the dependent variable leukoplakia. Pindborg et al (1967) in their study of oral leukoplakia in Lucknow pointed out that the habit of bidi smoking was associated with a far higher percentage of leukoplakia than any other smoking or chewing habits. In a follow-up study of Bombay policemen Mehta et al. (1969a) came to the same conclusion. The study by Pindborg et al. (1968a) in New Guinea also indicated that oral leukoplakia was more closely associated with smoking than with chewing. On the other hand Mehta et al. (1969b), in their epidemiological survey in four states of India, clearly showed that smoking and chewing habits may vary from one geographical area to another and that the significance of bidi smoking in comparison with that of other habits may change accordingly.

In the present study most of the influence of age upon presence of oral leukoplakia could be explained by differences in bidi smoking. However, age seemed to have some independent influence, which may be explained by the fact that oral leukoplakia is a chronic disease as demonstrated by Pindborg et al. (1968b). An increase in the number of leukoplakia patients with increasing age would thus be expected.

It is well known that oral leukoplakia is more prevalent among males than among females. It is therefore of interest that the present analysis showed that nearly all of the association between sex and leukoplakia could be explained by differences in bidi smoking. The remaining unexplained part of the association between oral leukoplakia and sex may be a result of the small numbers in some of the groups preventing more detailed statistical analysis. If this were so, studies of larger samples might show that the sex variable could be excluded once the tobacco habits could be analysed in greater detail.

It was found that the location of leukoplakia within the oral cavity was mainly determined by the habit-variable, as was found by Mehta et al. (1969b, 1969c). On the basis of the present data, however, a multi-variable influence of habit, sex, and age cannot be excluded.

The finding of a predilection of oral leukoplakia for the commissures and buccal mucosa is in agreement with earlier studies. Pindborg et al. (1967) reported that no cases of leukoplakia of the floor of the mouth were found in the study sample from Lucknow and similarly Mehta et al. (1969b) found only 1 case among 881 patients with leukoplakia in a survey covering 50 915 persons. In the present study 1 leukoplakia was located in this region. In contrast Roed-Petersen & Renstrup (1969) found 36 (6.4 %) leukoplakias of the floor of the mouth among 560 lesions in Danish patients.

Among patients with unilateral buccal leukoplakias a predilection for the right side was found. This may well be related to a side predilection in habits, but this possible association has not been further investigated.

Considering that the entropy estimated for leukoplakia in this study was found to be 0.132 it was surprising that the three independent variables bidi smoking, sex, and age explained no more than $\hat{I}$(leukoplakia; (bidi smoking, sex, age)) = 0.032 of this variation. This may mean that the presence of leukoplakia is determined by variables other than those included in the present study. This concept would be in agreement with the finding of Roed-Petersen, (1971) that the development of carcinoma in oral leukoplakias may not be determined only by those factors at present recognized as being involved in the etiology of oral leukoplakia.

RÉSUMÉ

ASSOCIATION ENTRE LA LEUCOPLASIE BUCCALE ET LE SEXE, L'ÂGE ET L'USAGE DU TABAC

Afin de déterminer la prévalence de la leucoplasie buccale et d'étudier les liens entre cette affection et certains facteurs étiologiques présumés, on a examiné une série continue de 9982 malades fréquentant la polyclinique du Dental College à Bombay (Inde), entre décembre 1963 et mai 1964. Avant l'examen, chaque patient a été interrogé sur son âge, sa religion et ses habitudes en matière de consommation de tabac à fumer ou à chiquer et d'usage de bétel.

Des lésions de leucoplasie buccale ont été découvertes chez 291 (2,9%) des sujets examinés. Ces cas ont fait l'objet d'une analyse statistique destinée à déceler une association éventuelle entre diverses variables indépendantes (sexe, âge, habitude de fumer ou de chiquer) et la maladie. L'usage de *bidis* (cigarettes bon marché à base de tabac indigène) apparaît comme le principal facteur étiologique. Pour une large part, l'influence de l'âge, et dans une mesure plus grande encore celle du sexe, peuvent être rapportées à des différences dans la consommation de tabac. Le rôle de l'âge n'est cependant pas négligeable étant donné que la leucoplasie buccale est une affection chronique et que les groupes d'âge élevé comptent un plus grand nombre de cas de cette affection. Quant au sexe, son influence s'estompe si l'on considère des sous-groupes formés de fumeurs et de non-fumeurs de *bidis*, mais elle réapparaît parmi les consommateurs d'autres préparations. Le rôle exact de ce facteur devrait être étudié sur des groupes de population plus importants. La localisation de la leucoplasie est, elle aussi, principalement déterminée par la variable « habitude de fumer ou de chiquer », mais ici une influence combinée de l'habitude, du sexe et de l'âge ne peut être exclue.

Selon les auteurs, l'intervention de facteurs étiologiques comme le sexe, l'âge ou l'usage du tabac ne suffit pas à expliquer complètement la présence ou l'absence de la leucoplasie buccale et d'autres éléments, encore inconnus, entrent certainement en ligne de compte.

REFERENCES

Larsen, M. E. (1968) [*Statistical analysis of elementary cybernetic systems*], Dissertation, Institute of Mathematical Statistics and Operational Analysis, Technical University of Denmark

Mehta, F. S. et al. (1969a) *Oral Surg.*, **28**, 372-388

Mehta, F. S. et al. (1969b) *Cancer*, **24**, 832-849

Mehta, F. S. et al. (1969c) *Brit. dent. J.*, **127**, 73-79

Pindborg, J. J. et al. (1965) *J. All-India dent. Ass.*, **37**, 228-229

Pindborg, J. J. (1967) *Bull. Wld Hlth Org.*, **37**, 109-116

Pindborg, J. J. et al. (1968a) *Cancer*, **22**, 379-384

Pindborg, J. J. et al. (1968b) *J. Amer. dent. Ass.*, **76**, 767-771

Roed-Petersen, B. & Renstrup, G. (1969) *Acta odont. scand.*, **27**, 681-695

Roed-Petersen, B. (1971) *J. dent. Res.*, **50**, 711.