Multivariate Statistical Modeling of the Factors Affecting Oral Health Disease - A Periodontal Disease

Dr. Javali S. B. Ph. D (shivalingappa.javali@gmail.com)
Department of Biostatistics, SDM College of Dental Sciences, Sattur

Modelare statistică multivariată a factorilor care afectează bolile de sănătate orală – boala periodontală

Boala periodontală, cea mai cunoscută boală orală care afectează omenirea şi ocupă un rol important în deciziile statutului sănătății orale în lume. Într-un studiu efectuat s-au depus eforturi pentru a determina factorii posibili care afectează boala periodontală şi a alege un model meticulous al bolii. Datele au fost culese de Indicele comunitar periodontal, pentru indicele de nevoi de tratament CPITN urmate de criteriile de diagnostic ale Organizației Mondiale a Sănătății (OMS) pe un eşantion sistematic aleatoriu din 1760 subiecți între 18-40 ani din Dharwad, Karnataka, India. Regresia logistică multiplă a fost estimată, fiind o abordare efectivă pentru răspunsuri binare față de modele cu influențe de profil la diverşi factori. Pentru a explora efectul combinat al fiecărui factor privind boala periodontală dicotomică din regresia logistică multiplă şi prin compararea performanțelor modelului logistic complet cu modelul logistic redus s-au folosit estimări logistice şi criterii de informații akaike. Valoarea criteriilor akaike a modelului redus este mai mică (12539) decât a modelului complet (12577). În concluzie, modelul regresiei logistice reduse este relativ mai bun decât modelul regresiei logistice complete la date de indici binari CPITN.

Cuvinte cheie: index CPITN, boala periodontală, modelul regresiei logistice

Abstract

Periodontal disease is the most common oral diseases that affect mankind and it occupies a prominent role in deciding the oral health status through out the world. In this study, an effort has been made to determine the most likely factors affecting periodontal disease and to select a meticulous model of the periodontal disease of study subjects. The data were collected by Community Periodontal Index for Treatment Needs (CPITN) index followed by WHO diagnostic criteria from a systematic random sample of 1760 subjects aged between 18-40 years in Dharwad, Karnataka, India. The Multiple Logistic Regression (MLR) was estimated; it is an effective approach for binary responses as compared with models for profiling influences of different factors. To explore the combined effect of each factor on dichotomous periodontal disease by MLR and compared the performances of full logistic model with that of reduced logistic model (step wise) using log likelihood estimate and Akaike Information Criterion (AIC). The AIC value of reduced model is smaller (1.2539) than that of full model (1.2577). It concluded that, the reduced logistic regression model is slightly a better fit as compared to full logistic regression model to the binary CPITN index data.

Key Words: CPITN Index, Periodontal disease, Logistic Regression Model (MLR), AIC

***

A healthy life is the dream of every individual irrespective of any physical or social differences; oral health is an integral part and very much important for the achievement & maintenance of general health. In developing countries like India, the present trend indicates that there is an increase in oral health problems especially gum disease. It occupies a prominent role in deciding the oral health status. Gum or periodontal disease remains a significant oral health problem and is a major cause of tooth loss in adults throughout the world. Despite the continuing scientific advances geared towards the treatment of periodontal disease, early diagnosis is essential to limit the extent and increase the potential for success of any definitive therapy provided. In addition, failure to diagnose and treat periodontal disease or provide timely referral of patients for treatment may lead to litigation (Davis, 1994 and Zinman, 2000). Numerous screening systems have been developed to detect periodontal disease. Some classic screening systems are: Periodontal Index (Russell, 1956), Periodontal Disease Index (Ramfjord, 1959), Community Periodontal Index of Treatment Needs (Ainamo et al., 1982) and the Extent and Severity Index (Carlos et al., 1986) etc. and all screening systems are effective. But in the present investigation, the Community Periodontal Index of Treatment Needs index is used for the assessment of periodontal disease status. This index is in ordinal scale and used to assess the periodontal disease. For simplicity and analytical purpose, the ordinal data is converted into dichotomous or binary (CPITN=0 and CPITN>0) outcome variable. Therefore, the greater attention has been paid to find out the factors, which are influencing periodontal disease. Therefore, the regression models have become an integral component of any data analysis concerned with the explanation of relationship between a response variable and one or more explanatory variables called covariates. Many different types of linear models have been seen in the literature and its use is discussed in many areas including dental epidemiology. The use of logistic regression modeling has been exploded during the past few decades. This method is now commonly applied in many fields including dental epidemiology. The logistic regression model is an important method to understand the principle that the goal of an analysis is the same as that of traditional model building technique used in statistical theory to find suitable description of relationship between response variable and a set of covariates. In traditional linear regression techniques we assumed that dependent variable must be continuous or quantitative.

In logistic model, we consider situations where the response variable is a categorical or ordinal random variable, attaining only two possible outcomes called binary or dichotomous. This difference between logistic and linear regression is reflected both in the choice of a parametric model and in the assumptions. In this article, the periodontal disease by Community Periodontal Index for Treatment Needs (CPITN) index is considered as dichotomized response variable and it is inappropriate to assume that they are normally distributed. Thus, the data cannot be analyzed using the traditional linear regression methods. It is convenient to denote one of the outcomes of response as without and with periodontal disease.

It is a standard practice to let the Y (Periodontal disease) to be two binary or dichotomous response variables, which are defined as $Y(Periodontal\ disease\ or\ CPITN\ Index) = \begin{cases} 1, & if\ Y\ (CPITN) > 0 \\ 0, & if\ Y\ (CPITN) = 0 \end{cases}$. The main aim and goal of this article is to modeling and utilization of multiple logistic regression models in identifying the important factors which are influencing on periodontal disease by CPITN index.

Methods and materials

Source of data and Study area

The cross sectional study was carried out to establish the significant factors of dichotomous periodontal disease among Dharwad city population, Karnataka state, India. To make more representative, a Dharwad city is divided into four zones (East zone, West zone, South zone and North zone) and then convenient sample of two wards were selected randomly from each zone. From selected convenient of eight wards, the random samples of 600 households were selected (75 households from each zone). Lastly, systematic samples of 1760 individuals aged between 18-40 years were included.

Clinical examination of periodontal disease

The data on periodontal disease was collected by clinical examination, which was carried out by two qualified dentists using CPITN index (Community Periodontal Index for Treatment Needs index) followed by criteria recommended by WHO (WHO, 1997) with plane mouth mirror, dental explorer, disposable gloves and sterilized instruments under artificial light. Before the start of the actual study, a pilot study was conducted on convenient sample of 50 subjects to assess the intra and inter examiner agreement for recording CPITN index scores. The intra-examiner and inter examiner agreement found to be respectively 0.8619 and 0.9018.

Response Variable and Independent factors

Periodontal disease i.e. CPITN index is an ordinal variable which is taken as the response variable. For analysis purpose, the CPITN index data were grouped as 0 if CPITN =0 and 1 if CPITN >0 and deliberated here as dichotomous response variable. Apart from response variable, the data set on independent factors like socio-economic-demographic, food habits, eating habits, oral hygiene practices and deleterious habits obtained by structured questionnaire and interview method (Table 1) and all independent variables are binary or dichotomous except age is considered as continuous variable. The information on these independent variables was collected by structured questionnaire with personal interview method.

Code sheet for the independent factors

Table 1

No	Description	Code/values
1	Gender	Male=0, Female=1
2	Age (in years)	As a continuous
3	Religion	Hindu=1, Non-Hindu =2
4	Caste	SC/ST/OBC=1, GM=2
5	Socio-economic status	Low=0, High=1
6	Family size	≤5 =0, >5=1
7	Staple food	Wheat/Rice/Jower=1, Others=2
8	Sources of drinking water	Pipeline/River/Pound=1, Tube well/Hand pump=2
9	Types of diet	Vegetarian=0, Non-vegetarian=1
10	Timings of sweet consumption	During/Between meals=0, During and between meals=1
11	Frequency of sweet consumption	≤2 times =0, >2 times =1
12	Oral hygiene habits	Finger/datun/others =0, Tooth Brush =1
13	Frequency of brushing	Once=1, Twice or more=2
14	Timings of cleaning the teeth	Morning or Night=1, Morning and Night=0
15	Methods of brushing	Circular/ Vertical=1, Horizontal=2
16	Materials used for brushing their teeth	Paste/powder=1, Others=2
17	Types of toothpaste	Non-fluoridated=0, Fluoridated=1
18	Mouth rinsing habit	No=0, Yes=1
19	Smoking habit	No=0, Yes=1
20	Chewing habit	No=0, Yes=1
21	Alcohol habit	No=0, Yes=1

Data Analysis

The dichotomized periodontal disease data are analyzed and multiple logistic regression model is constructed between the binary response variable with independent variables. The model estimation, in the first step, the multiple logistic full model is constructed for considering all independent variables and in the second step, the stepwise called multiple logistic reduced model is performed by considering only significant variables from the full model. In the selection procedure using the stepwise multiple logistic model analysis, we first select the variable having a greatest influence power. Then the effect of this variable is eliminated from the information content of all the other variables. The variable, which then has the greatest power of influence after the above elimination procedure, is ranked as the second etc. Thus, the variables are listed in decreasing order with respect to the probability of additional information gained from including further variables was less than 0.05. The variable having the weaker influence power may be dropped from the final analysis. In order to weigh the significance of each chosen variable with respect to their influence, its correlation with multiple logistic regression model and parameter estimates, standard error of estimates, Odds Ratios (OR), 95% confidence intervals (95% CI) and p-values of each *variable computed. Also the fitting performance of full and reduced multiple full logistic regression models evaluated on the basis of Log likelihood statistic and Akaike's information criterion (AIC) statistic (*Akaike H. *1974). A statistical significance was set at 5% level of significance (p<0.05).*

Formulation of Multiple Logistic Models

Consider a collection of independent variables (atleast interval scale) denoted by the vector $X' = (x_1, x_2, ..., x_p)$. Let the conditional probability that the response variable is present be denoted by p[Y=1 | x] = π (x). The logit of the multiple regression model is given by the equ ation

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ...\beta_p x_p .$$

with the logistic model given by

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

If some of the variables such as gender, socio-economic status etc. are discrete that are measured in nominal scale and so forth is inappropriate to include them in the model unless if they are interval scale variables. The number used to represent the various levels of these nominal scale variables are merely identifiers and have no numeric significance. In this situation the method of choice is to use a collection of design variables (or dummy variables).

In general, if a nominal scaled variable has k possible values, then k-1 design variables are needed. Thus, the logit for a model with p variables and the j^{th} variable being discrete would be

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \sum_{l=1}^{k_j - 1} \beta_{jl} D_{jl} + \beta_p x_p$$

While discussing the multiple logistic regression models, in general suppress the summation sign and when design variables are being used.

Fitting the Multiple Logistic Regression Model

Assuming a sample of n independent observations (x_i, y_i) i=1, 2, 3… n., fitting the model requires estimates of the vector $\beta' = (\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$. The likelihood of β is given by

$$l(\beta)=\prod_{i=1}^{n}\pi(x_i)^{y_i}[1-\pi(x_i)]^{1-y_i}\text{, where } \pi(x)=\frac{e^{g(x)}}{1+e^{g(x)}}$$

The likelihood function of equation (7.3) is given by

$$\log\ l(\beta)=L(\beta)=\log[\prod_{i=1}^{n}\pi(x_i)^{y_i}[1-\pi(x_i)^{1-y_i}]$$

$$=\sum_{i=1}^{n}\{y_i\log[\pi(x_i)]+(1-y_i)\log[1-\pi(x_i)]\}$$

Here we get (p+1) likelihood equations that are obtained by differentiating the log likelihood function with respect to the p+1 coefficient. The likelihood equations obtained may be expressed as follows:

$$\sum_{i=1}^{n}[y_i-\pi(x_i)]=0$$

and

$$\sum_{i=1}^{n}x_{ij}[y_i-\pi(x_i)]=0 \quad \text{for j=1, 2, …, p.}$$

Let $\widehat{\beta}$ denote the solution to these equations. Thus, the fitted values of the multiple logistic regression model are $\widehat{\pi}(x_i)$. Then, the method of estimating the variance and co-variances of these estimated coefficients follows from the well developed theory of maximum likelihood estimation (Rao 1973, David and Stanley 2000). This theory states that estimators are obtained from the matrix of partial derivatives of the log likelihood function. These partial derivatives have the following general form:

$$\frac{\partial^2 L(\beta)}{\partial\beta_j^2}=-\sum_{i=1}^{n}x_{ij}^2\pi_i(1-\pi_i)\text{ and}$$

$$\frac{\partial^2 L(\beta)}{\partial\beta_j\partial\beta i_j}=-\sum_{i=1}^{n}x_{ij}x_{il}\pi_i(1-\pi_i)\text{, for j, l=0, 1, 2… p}$$

and $\pi_i=\pi_i(x)$

Further, the estimated standard errors of the estimated coefficients of the logistic regression model is given by

$$S\hat{E}(\hat{\beta}_j) = \sqrt{V\hat{a}r(\hat{\beta}_j)} \text{ or } \left[V\hat{a}r(\hat{\beta}_j)\right]^{1/2} \text{ for j = 0, 1, 2, …,p.}$$

Alternatively, the Wald or Z statistic is commonly used to test the significance of individual logistic regression coefficients for each independent variable. The test statistic is given by

$$W_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}, j = 0,1,2,..., p$$

The multivariable analog of the Wald test is given by

$$N = \hat{\beta}_j \left[V\hat{a}r(\hat{\beta})\right]^{-1} \hat{\beta} = \hat{\beta}^I (XVX)\hat{\beta}$$

The statistic N is distributed as chi-square with [p+1] degrees of freedom under the hypothesis that each of the p+1 coefficient is equal to zero. The multivariable analog of the score for the significance of the model is based on the distribution of the p derivatives of L (β) with respect to β.

Further, the sensitivity; specificity are used to for determining the presence and absence of a disease. Also, the receiver operating characteristic (ROC) curve analysis is used to diagnostic performance of a test or the accuracy of a test to discriminate diseased cases from normal cases is evaluated (Metz, 1978; Zweig & Campbell, 1993).

Results

A total of 1760 subjects are included in the study (50.00% are males and 50.00% are females) which has mean age as 34.26 and mean family size as 2.94. Similarly, 1200 (68.18%) are Hindus, 1687 (95.85%) are backward castes, 1025 (58.24%) are with high socio-economic status, 292 (16.59%) are users of wheat or rice or jower as a main staple food, 1432 (81.36%) are drinking Tube well / hand pump water, 1002 (56.93%) are non-vegetarian, 1644 (93.41%) are eating sweet in during or between meals, 1674 (95.11%) are taking sweet consumption at least twice in a day, 929 (52.78%) are brushing their teeth with tooth brushes as a oral hygiene habit, 1466 (83.30%)

are brushing their teeth with only once in a day, 1425 (80.97%) are brushing their teeth in the both morning and night, 1461 (83.01%) are brushing their teeth by horizontal method, 1273 (72.33%) are brushing their teeth by paste/powder, 1150 (65.34%) are users of non-fluoridated toothpastes, 962 (54.66%) are changing their toothbrush once in after four months, 1154 (65.57%) are not rinsing their mouth after every meal with water, 1352 (76.82%) are smokers, 840 (47.73%) are chewers and 962 (54.66%) are alcohol drinkers as compared to their counterparts. Table 2 presents parameter estimates and their standard errors of covariates of periodontal disease (CPITN Index) using multiple logistic regression model. A total of 21 covariates are included in the model, in which only 5 covariates are found to be significant ($p<0.05$). Among significant covariates, only one covariate such as family size has positive association with periodontal disease. The regression coefficient corresponding to significant covariate is found to be positive. However, four covariates namely gender, frequency of brushing, timings of cleaning the teeth and type of toothpastes are negatively associated with periodontal disease. These significant covariates exhibited negative regression coefficients.

Further, log likelihood of this model is -1085.7876. The Akaike's Information Criterion (AIC) value is 1.2577. Our goal here is to estimate the best fitting model of periodontal disease while minimizing the number of covariates. The next logical step is to fit a reduced regression model containing only those significant covariates and comparing it with the full regression model containing all 21 covariates. The results of the fitted reduced regression model with estimated coefficients, p-value and log likelihood and AIC are presented in Table 3. The reduced regression model is obtained by removing the insignificant covariates from the full regression model. The log likelihood and AIC value of the reduced regression model are -1098.4320 and 1.2539 respectively. Based on log likelihood and AIC values, the full and reduced logistic regression models are similar. Thus, there is no advantage in excluding some covariates from the model for assessment of significant determinants of occurrence of periodontal disease.

The Estimated Coefficients of Covariates from Full Logistic Regression Model to Periodontal Disease Dichotomous Data

Table 2

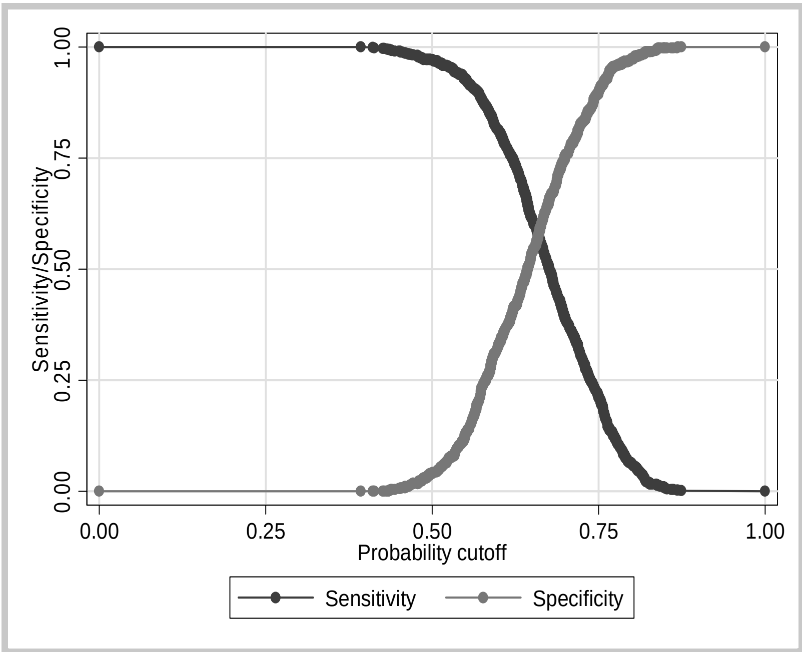
Covariates	Estimate	Std. Err.	z-value	Confidence interval	
				+96%	-95%
Constant	1.7142	0.8427	2.0300	0.0626	3.3658
Gender	0.4064	0.1168	3.4800*	0.1775	0.6354
Age (in years)	0.0531	0.0391	1.3600	-0.0235	0.1297
Religion	-0.0246	0.1125	-0.2200	-0.2451	0.1959
Caste	-0.0619	0.0644	-0.9600	-0.1881	0.0642
Socio-economic status	-0.1165	0.0827	-1.4100	-0.2787	0.0457
Family size	0.2981	0.1598	1.8700*	-0.0150	0.6113
Staple food	0.0481	0.1826	0.2600	-0.3099	0.4060
Sources of drinking water	-0.3023	0.1990	-1.5200	-0.6923	0.0877
Dietary habits	-0.1461	0.1078	-1.3600	-0.3574	0.0651
Time for sweet consumption	0.0978	0.3221	0.3000	-0.5336	0.7292
Frequency of sweet consumption	0.3592	0.3746	0.9600	-0.3751	1.0935
Oral hygiene habits	-0.0722	0.1046	-0.6900	-0.2772	0.1327
Frequency of brushing	-0.3533	0.1430	-2.4700*	-0.6335	-0.0730
Timings of cleaning the teeth	-0.3069	0.1431	-2.1400*	-0.5874	-0.0264
Methods of brushing	-0.0174	0.1189	-0.1500	-0.2504	0.2156
Materials used for brushing their teeth	0.1733	0.1684	1.0300	-0.1567	0.5034
Type of toothpastes	-0.4708	0.1418	-3.3200*	-0.7486	-0.1929
Mouth rinsing habit	0.0093	0.1173	0.0800	-0.2205	0.2392
Smoking habit	0.0199	0.1325	0.1500	-0.2398	0.2795
Chewing habit	0.2154	0.2095	1.0300	-0.1953	0.6260
Alcohol habit	0.3237	0.2130	1.5200	-0.0937	0.7411
Log likelihood	-1085.7876				
AIC	1.2577				

*significant at 5%level of significance (p<0.05)

The plot of sensitivity and specificity versus criterion value for the response variable (CPITN Index) in the full and reduced regression model is presented respectively in Figure 1 and Figure 3.

The Plot of Sensitivity and Specificity versus Criterion Value for the Response Variable (CPITN Index) in the Full Model

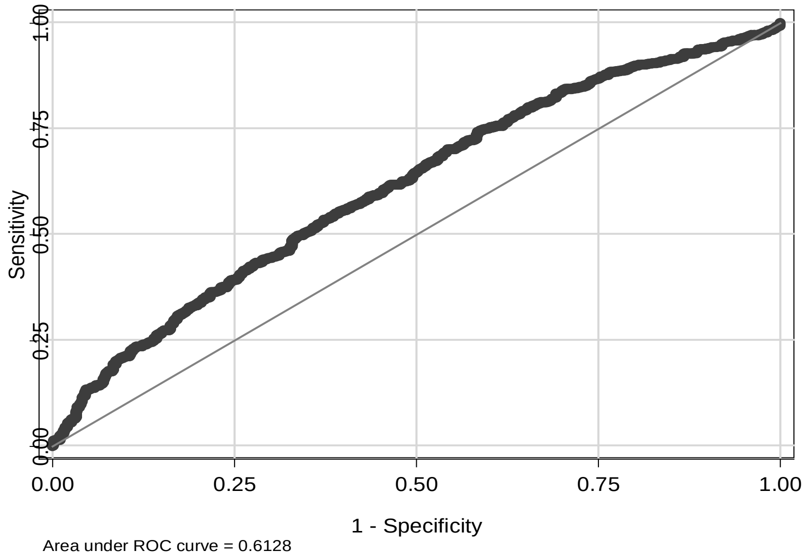
Figure 1



The area under Receiver Operating Characteristic (ROC) curve of the response variable (CPITN Index) for the full model is 0.6128 and 0.5821 is in reduced model. It provides a summary of the accuracy of the diagnostic test which is nearly respectively 61% and 58% for full and reduced models (Figure 2 and Figure 4).

The Accuracy of the Test in the Means of ROC (CPITN Index) for Full Model

Figure 2



The Estimated Coefficients of Covariates from Step Wise Logistic Regression Model to Periodontal Disease Dichotomous Data

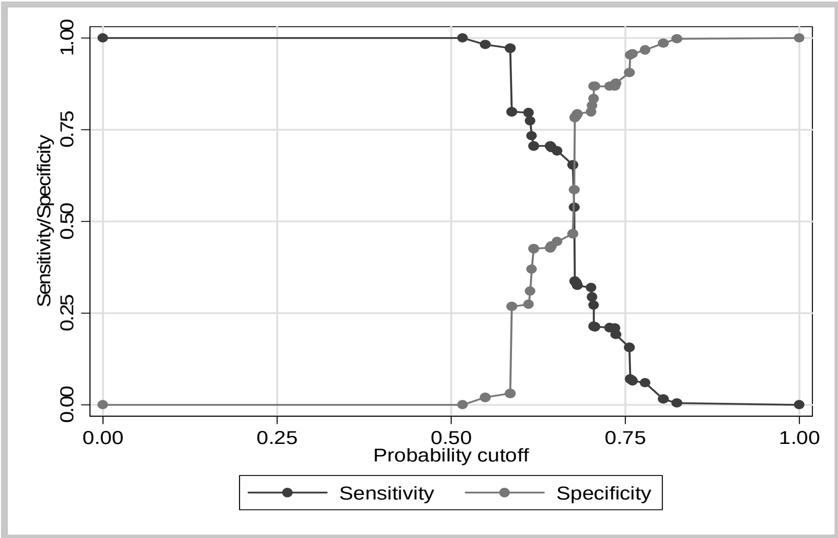
Table 3

Covariates	Estimate	Std. Err.	Z-value	Confidence interval	
				-95%	+95%
Constant	1.1382	0.2185	5.2100*	0.7100	1.5665
Gender	0.3980	0.1035	3.8500*	0.1951	0.6009
Family size	0.2881	0.1579	1.8200*	-0.0215	0.5976
Frequency of brushing	-0.2769	0.1368	-2.0200*	-0.5451	-0.0088
Type of toothpastes	-0.3888	0.1110	-3.5000*	-0.6064	-0.1712
Log likelihood	-1098.4320				
AIC	1.2539				

*significant at 5%level of significance (p<0.05)

The Plot of Sensitivity and Specificity versus Criterion Value for the Response Variable (CPITN Index) in the Reduced Model

Figure 3



The Accuracy of the Test in the Means of ROC (CPITN Index) for Reduced Model

Figure 4

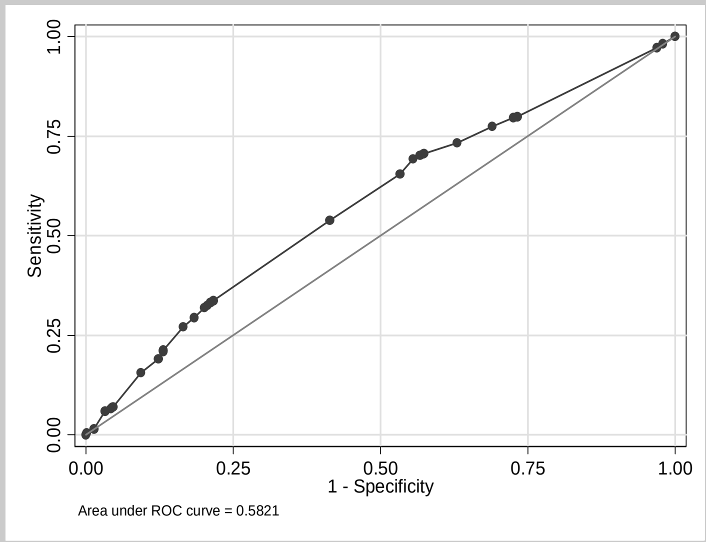


Table 4 represents the odds ratio, standard error, 95% confidence intervals of selected covariates on periodontal disease by full logistic regression model. It reveals that, the estimated odds ratio of gender (OR=1.5015, 95% CI: 1.1942-1.8878), family size (OR=1.3474, 95% CI: 0.9851-1.8428), frequency of brushing (OR=0.7024, 95% CI: 0.5307-0.9296), timings of cleaning the teeth (OR=0.7357, 95% CI: 0.5558-0.9740) and type of toothpastes (OR=0.6245, 95% CI: 0.4730-0.8246) have found to be significant ($p<0.05$). It means that, the gender, family size, frequency of brushing, timings of cleaning the teeth and type of toothpastes have a significant impact on periodontal disease. In other words, the women living in a larger family (>5 members in a family), brushing their teeth only once a day, who are brushing their teeth morning and night without pastes/powder have a significant higher prevalence of periodontal disease as compared to their counterparts.

However, the odds ratio, standard error, 95% confidence intervals of covariates on periodontal disease in reduced model has been presented in Table 5. It reveals that there is an improvement in the strength of association among some covariates.

The Estimated Odds Ratios of Covariates from Full Logistic Regression Model to Periodontal Disease Dichotomous Data

Table 4

Covariates	Odds ratio	Std. Err.	Z-value	Confidence interval	
				+96%	-95%
Gender	1.5015	0.1754	3.4800*	1.1942	1.8878
Age (in years)	1.0545	0.0412	1.3600	0.9767	1.1385
Religion	0.9757	0.1098	-0.2200	0.7826	1.2164
Caste	0.9399	0.0605	-0.9600	0.8285	1.0663
Socio-economic status	0.8900	0.0736	-1.4100	0.7568	1.0467
Family size	1.3474	0.2153	1.8700*	0.9851	1.8428
Staple food	1.0492	0.1916	0.2600	0.7336	1.5008
Sources of drinking water	0.7391	0.1471	-1.5200	0.5004	1.0916
Dietary habits	0.8640	0.0931	-1.3600	0.6995	1.0673
Time for sweet consumption	1.1027	0.3552	0.3000	0.5865	2.0734
Frequency of sweet consumption	1.4321	0.5365	0.9600	0.6872	2.9846
Oral hygiene habits	0.9303	0.0973	-0.6900	0.7579	1.1420
Frequency of brushing	0.7024	0.1004	-2.4700*	0.5307	0.9296
Timings of cleaning the teeth	0.7357	0.1053	-2.1400*	0.5558	0.9740
Methods of brushing	0.9827	0.1168	-0.1500	0.7785	1.2406
Materials used for brushing their teeth	1.1893	0.2003	1.0300	0.8550	1.6543
Type of toothpastes	0.6245	0.0886	-3.3200*	0.4730	0.8246
Mouth rinsing habit	1.0093	0.1184	0.0800	0.8021	1.2702
Smoking habit	1.0201	0.1351	0.1500	0.7868	1.3225
Chewing habit	1.2403	0.2599	1.0300	0.8226	1.8702
Alcohol habit	1.3822	0.2944	1.5200	0.9105	2.0982

*significant at 5%level of significance ($p<0.05$)

The Estimated Odds Ratios of Covariates from Step Wise Logistic Regression Model to Periodontal Disease Dichotomous Data

Table 5

Covariates	Odds ratio	Std. Err.	Z-value	Confidence interval	
				-95%	+95%
Gender	1.4889	0.1541	3.8500*	1.2155	1.8238
Family size	1.3339	0.2107	1.8200*	0.9788	1.8178
Frequency of brushing	0.7581	0.1037	-2.0200*	0.5798	0.9912
Type of toothpastes	0.6779	0.0753	-3.5000*	0.5453	0.8427

*significant at 5%level of significance (p<0.05)

Discussion and Conclusions

Changes in our knowledge of the etiology of periodontal disease and the recognition of the potential importance of susceptibility factors as they affect initiation and progression of periodontal disease, have led to an intense study of specific risk factors for periodontal disease. The gender factor is associated with periodontal disease. It means that, the periodontal disease is more prevalent in males than in females at any comparable ages. This result coincides with several studies done by Miller et al., 1987 and Grossi et al., 1994 and 1995. Males usually exhibit proper oral hygiene than females (U.S. Public Health Service, 1979). The reasons for these gender differences are not clear and their elucidation may reveal important destructive or protective mechanism.

The age is an insignificant factor having positive association with periodontal disease in the study. However, the studies on periodontal disease prevalence with extent and severity show that disease is more prevalent in older age groups as compared to younger groups (Miller et al., 1987; Grossi et al., 1994, 1995; Marshal et al., 1955; Schei et al., 1959 and Abdellatif et al., 1987). Also it is found that the sevierity of the disease is more with respect to plaque development and gingivitis in elderly persons as compared to younger persons (Abdellatif et al., 1987).

The relationship of periodontal disease and socioeconomic status can be viewed globally, where wide variations in socio-economic status among different populations are compared. These studies compare populations from developing countries with those from industrialized countries which suggest that periodontal disease may be associated with nutritional deficiencies (Russell, 1960). However, in this study an association is not found to be statistically significant. But, the Ramfjord et al. (1968) found that the periodontal condition of young men in India who exhibited clinical symptoms of general malnutrition is not different from that of the periodontal condition of well nourished individuals. Non-Hindus showed that they have apparently more periodontal destruction compared to Hindus. No studies are found in relation to religion on Indian population with respect to periodontal disease.

There is a history of association between tobacco smoking and periodontal disease (Pindborg, 1947; Frandsen and Pindborg, 1949; Solomon, Priore and Bross, 1968) and prevalence of Acute Ulcerative Gingivitis (ANUG) was demonstrated as early in 1946 (Pindborg, 1947 and 1949). However, the perception that greater levels of plaque and calculus is more in smokers than that in non-smokers. In this study, it is shown that smoking tobacco is not significantly associated with periodontal disease. This result coincides with some of the earlier studies (Bergstrom and Floderus Myrhed, 1983; Preber et al, 1980; Bergstrom, 1981, 1990; Sheiham, 1971; Macgregor, Edgar and Greenwood, 1985; Preber and Bergstrom, 1986; Schei, Waerhaug, Lovdal et al., 1959; Herulf, 1968; Solomon, Priore and Bross, 1968). It is likely that smoking is a major factor for destructive periodontal disease in man. Hence the modification of this factor is important in the treatment and prevention of periodontal disease.

Further, in this article, we compared performance of full logistic model with that of reduced logistic model using log likelihood estimate of CPITN index data. The results show that the fitting performance of full logistic regression model is slightly better as compared to reduced logistic regression model applied to dichotomized CPITN index data.

Bibliography

- Davis L. G. (1994). The effects of porcelain laminate veneers on gingival health and bacterial plaque characteristics. Journal of Clinical Periodontology, Vol. 21(9), PP. 638-640.
- Zinman (2000). Dental and legal considerations in periodontal therapy. Periodontology, Vol. 25(1), PP. 114-130.
- Russell A. L. (1956). A system of classification and scoring for prevalence surveys of periodontal disease. J. Dent. Res., Vol. 35, P. 350.
- Ramfjord S. P. (1959).Indices for prevalence and incidence of periodontal disease. J. Periodontology, Vol. 30, P. 51.
- Ainamo J., Barnes D., Beagrie G. et al. (1982). Development of the World Health Organization (WHO) community periodontal index of treatment needs (CPITN). Int. Dent. J., Vol. 32(3), PP. 281-291.

- Carlos J. P., Wolfe M. D. and Fingman A. (1986). The extent and severity index: a simple method for use in epidemiologic studies of periodontal disease. J. Clin. Periodontology, Vol. 13(5), PP. 500-505.

- World Health Organization (1997). Oral health surveys. Basic Methods. WHO Geneva.

- Akaike H. (1974). "A new look at the statistical model identification". IEEE Transactions on Automatic Control, Vol.6, PP.716–723,

- Rao C. R. (1973). "Linear Statistical Inference and its application", Second Edition, Wiley Inc., New York.

- David W. H. and Stanley (2000). "Applied Logistic Regression", Second edition, A Wiley Inter science Publication, John Wiley and Sons, Inc. New York.

- Metz C. E. (1978). "Basic principles of ROC analysis". Seminars in Nuclear Medicine, Vol.8, PP. 283-298.

- Zweig M.H. and Campbell G. (1993). "Receiver-operating characteristic (ROC) plots, a fundamental evaluation tool in clinical medicine". Clinical Chemistry, Vol.39, PP.561-577.

- Miller A. J., Brunelle J. A., Carlos J. P., Brown L. J. and Loe H. (1987). Oral Health of United States Adults. National findings. Bethesda MD. National Institute of Dental Research. NIH publication No. 87, P. 28-68.

- Grossi S. G., Zambon J. J., Ho A. W. et al. (1994). Assessment of risk for periodontal disease-I, Risk indicators for attachment loss. J. Periodontology, Vol. 65, PP. 260-267.

- Grossi S. G., Genco R. J., Machatei E. E. et al. (1995). Assessment of risk for periodontal disease. II. Risk indicators for alveolar bone loss. J. Periodontology, Vol. 66, PP. 23-29.

- U. S. Public Health Service, National center for Health Statistics (1979). Basic data on Dental examination findings of persons 1-75 years, United States 1971-1974. DHEW publication No. (PHS) 79-1662, Series 11. No. 214. Washington DC. Government printing office.

- Marshal-Day C. D., Stephens R. G. and Quigley L. F. Jr. (1955). Periodontal disease prevalence and incidence. J. Periodontology, Vol. 26, PP. 185-203.

- Schei O., Waerhaug J., Lovdal A. and Arno A. (1959). Alveolar bone loss as related to oral hygiene and age. J. Periodontology, Vol. 30, PP. 7-16.

- Abdellatif H. M. and Burt B.A. (1987). An epidemiological investigation into the relative importance of age and oral hygiene status as determinants of periodontitis. J. Dent. Res., Vol. 66, PP. 13-18.

- Russell A. L. (1960). Geographical Distribution and Epidemiology of Periodontal Disease. Geneva. World Health Organization, (WHO/DH/33/34)

- Ramfjord S. P., Emslie R. D. and Waerhaug J. (1968). Epidemiological studies of periodontal disease. Am. J. Public Health, Vol. 28, PP. 1713-1722.

- Pindborg J. J. (1947). Tobacco and Gingivitis I. Statistical examination of significance of tobacco in the development of ulcer membranous gingivitis and in the formation of calculus. J. Dent. Res., Vol. 26, PP. 261-264.

- Frandsen A. and Pindborg J. J. (1949). Tobacco and gingivitis III. Difference in action of cigarette and pipe smoking. J. Dent. Res. Dent., Vol. 28, PP. 464-465

- Solomon H. A., Priore R. L. and Bross I. D. (1968). Cigarette smoking and periodontal disease. J. Am. Dent. Assoc., Vol. 77, pp. 1081-1084.

- Pindborg J. J. (1949). Tobacco and Gingivitis II. Correlation between consumption of tobacco, ulcer membranous gingivitis and calculus. J. Dent. Res., Vol. 28, PP. 460-463.

- Bergstrom J. and Floderus-Myred B. (1983). Co-twin control study of the relationship between smoking and some periodontal disease factors. Community Dent. Oral Epidemiol., Vol. 11, PP. 113-116.

- Preber H., Kant T. and Bergstrom J. (1980). Cigarette smoking, oral hygiene and periodontal health in Swedish army conscripts. J Clin. Periodontology, Vol. 7. PP. 106-113.
- Bergstrom J. (1981). Short term investigation on the influence of cigarette smoking upon plaque accumulation. Scand. J. Dent. Res., Vol. 89, PP. 235-238.
- Bergstrom J. (1990). Oral hygiene compliance and gingivitis expression in cigarette smokers. Scand. J. Dent. Res., Vol. 98, PP. 497-503.
- Sheiham A. (1971). Periodontal disease and oral cleanliness in tobacco smokers. J. Periodontology, Vol. 42, PP. 259-263.
- Macgregor I. D. M., Edgar W. M. and Greenhood A. R. (1985). Effects of cigarette smoking on the rate of plaque formation. J. Clin. Periodontology, Vol. 12, PP. 35-41.
- Preber H. and Bergstrom J. (1986). Cigarette smoking in patients referred for periodontal treatment. Scand. J. Dent. Res., Vol. 94, PP. 102-108.