
Multiple logistic regression model to predict risk factors of oral health diseases

Dr. Shivalingappa B. Javali, Ph.D. (javalimanju@rediffmail.com)

SDM College of Dental Sciences, Karnataka, India

Dr. Parameshwar V. Pandit, Ph.D. (panditpv@rediffmail.com)

Bangalore University, Karnataka, India

Pentru a analiza dependența cariilor dentare și a bolilor parodontale, cu luare în considerare a numărului de factori de risc s-au avut în vedere cererile Modelului de regresie logistică. Studiul secțiunii transversale a implicat un eșantion aleatoriu sistematic din 1.760 de pacienți, cu vârste cuprinse între 18 și 40 de ani în Dharwad, Karnataka, India. Factorii de risc ai **cariilor dentare** și a **bolii parodontale** au fost stabiliți prin **Modelul de regresie multiplă** folosind software statistic. Frecvența de periaj, temporizările **de curățare a dinților** și **tipul de pastă de dinți** reprezintă principalii factori de risc semnificativi ai **cariilor dentare** și **bolilor parodontale**. Implementarea performanței Modelului de regresie logistic este redusă de o potrivire mai bună față de Modelul de regresie logistică completă în identificarea factorilor de risc pentru ambele situații.

Abstract

Purpose: To analysis the dependence of oral health diseases i.e. dental caries and periodontal disease on considering the number of risk factors through the applications of logistic regression model. **Method:** The cross sectional study involves a systematic random sample of 1760 permanent dentition aged between 18-40 years in Dharwad, Karnataka, India. Dharwad is situated in North Karnataka. The mean age was 34.26 ± 7.28 . The risk factors of dental caries and periodontal disease were established by multiple logistic regression model using SPSS statistical software. **Results:** The factors like frequency of brushing, timings of cleaning teeth and type of toothpastes are significant persistent predictors of dental caries and periodontal disease. The log likelihood value of full model is -1013.1364 and Akaike's Information Criterion (AIC) is 1.1752 as compared to reduced regression model are -1019.8106 and 1.1748 respectively for dental caries. But, the log likelihood value of full model is -1085.7876 and AIC is 1.2577 followed by reduced

regression model are -1019.8106 and 1.1748 respectively for periodontal disease. The area under Receiver Operating Characteristic (ROC) curve for the dental caries is 0.7509 (full model) and 0.7447 (reduced model); the ROC for the periodontal disease is 0.6128 (full model) and 0.5821 (reduced model). **Conclusions:** The frequency of brushing, timings of cleaning teeth and type of toothpastes are main significant risk factors of dental caries and periodontal disease. The fitting performance of reduced logistic regression model is slightly a better fit as compared to full logistic regression model in identifying the these risk factors for both dichotomous dental caries and periodontal disease.

Key Words: Dental caries, Periodontal disease, Akaike Information Criterion, Receiver Operating Characteristic, Full model, Reduced model

Dental caries and periodontal disease are remains a significant oral health problem and are a major cause of tooth loss in adults throughout the world. Despite the continuing scientific advances geared towards the treatment of these diseases, early diagnosis is essential to limit the extent and increase the potential for success of any definitive therapy provided. In developing countries like India, the present trend indicates that there is an increase in these diseases and it plays an important role in deciding the oral health status. These are multi-factorial diseases; the larger concentration has been paid to find out the significant risk factors, which are influencing dental caries and periodontal disease. Thus, the risk assessment has become increasingly important in the prevention of caries and periodontal disease. However, the changes in our knowledge of the etiology and the reorganization of the probable significance of susceptibility factors as they affect initiation and progression, have led to intense study of specific risk factors.

To investigate the risk factors associated with a dental caries and periodontal disease, the regression methods 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 factors. 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 biomedical research, business and finance, criminology, economics, ecology, engineering, health policy, medicine, agriculture and dental epidemiology, etc. 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 factors. In traditional linear regression techniques we assumed that dependent variable must be continuous or quantitative.

In logistic model, we consider the response variable is a categorical 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 response variables are dichotomous, 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 dental caries (or without and with periodontal disease).

Response Variables and Independent Factors

The Dental caries by DMFT Index and Periodontal Disease by CPITN Index are the two independent response variables. For convenience of statistical modeling, the DMFT index and CPITN index scores were grouped as 0 and 1 (DMFT=0, DMFT>0=1 and CPITN=0, CPITN>0=1) called binary response variable. Apart from response variables, the data set on independent factors like socio-economic-demographic, food habits, eating habits, oral hygiene practices and deleterious habits are obtained by structured questionnaire with personal interview method. However, all independent factors are binary in nature except age and socio-economic status. They are Gender, age (in years), religion, caste, socio-economic status, family size, staple food, sources of drinking water, types of diet, time for sweet consumption, frequency of sweet consumption, oral hygiene habits, frequency of brushing, timings of cleaning the teeth, methods of brushing, materials used for brushing their teeth, types of toothpaste, mouth rinsing habit, smoking habit chewing habit and alcohol habit. The data on these factors were collected and recorded by structured questionnaire with interview method.

Method and Materials

Study area, Study population and sampling procedure

The cross sectional study was conducted in Dharwad, Karnataka, India. Dharwad is situated in North Karnataka and is one of the educational centers. The cross sectional study involves a systematic random sample of 1760 permanent dentition aged between 18-40 years. Sample size was determined based on results of pilot study, which showed that SD of CPITN scores was 0.8120 under precision of 5% and 99% of confidence level, the sample size was estimated to be $1756 \square 1760$. The mean age of the study subject was 34.26 ± 7.28 .

Clinical Examination

The selected subjects were called for free dental caries and periodontal disease examination by two well qualified community dentists under artificial light for each subject. The findings of the dental caries by DMFT index and periodontal disease by CPITN index were recorded according to diagnostic criteria recommended by World Health Organization (1-2). A pilot study was conducted with a convenient sample of 150 to assess the intra and inter examiner agreement for DMFT and CPITN index scores. The intra-examiner agreement was 0.8719 (first examiner) and 0.7193 (second examiner) respectively for DMFT index and 0.7984 and 0.8710 for CPITN index. The inter examiner agreement was calculated by Kappa statistic and it was ranged from 0.6514 to 0.8795 for both indices.

Formulation of Multiple Logistic Regression Model

Nelder and Wedderburn (3) introduced generalized linear models which are a class of statistical models. It includes response variables that follow any probability distribution in the exponential family of distributions. An excellent treatment of generalized linear models is presented in Agresti (4). In this study the response variables are binary; therefore, the logistic regression model is an appropriate model, which is a part of generalized linear models. The response variable can take the value 1 with a probability of success P, or the value 0 with probability of failure 1-P. This type of variable is called a Bernoulli (or binary) variable. The relationship between the predictor and response variables is not a linear function in logistic regression; instead, the logistic regression function is used, which is given as

$$P(x) = \frac{e^{(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}}{1 + e^{(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}} \quad \text{We now find the link function for}$$

which the logistic regression model is a generalized linear model (GLM). For

$$\text{this model the odds of making response 1 are } \frac{P(x)}{1 - P(x)} = e^{(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)},$$

$$\log \left[\frac{P(x)}{1 - P(x)} \right] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k. \quad \text{Thus the appropriate link is the}$$

log odds transformation and the logit. The logistic regression model is given

$$\text{by } \text{it } [P(x)] = \frac{P(x)}{1 - P(x)} = e^{(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}. \quad \text{The parameters in this}$$

model, $\alpha, \beta_1, \beta_2, \dots, \beta_k$ can no longer be estimated by least squares, but are found using the maximum likelihood method (5). Logistic regression calculates the

probability of success over the probability of failure; therefore, the results of the analysis are in the form of an odds ratio. Logistic regression also provides knowledge of the relationships and strengths among the variables.

Fitting the Multiple Logistic Regression Model

Suppose assuming a sample of n independent observations (x_i, y_i) $i=1, 2, 3 \dots n$, fitting the model requires estimates of the vector

$\beta' = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)$. The likelihood of β is given by $l(\beta) = \prod_{i=1}^n P(x_i)^{y_i} [1 - P(x_i)]^{1-y_i}$, where $P(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$. The likelihood function is given by

$\log l(\beta) = \sum_{i=1}^n \{y_i \log[P(x_i)] + (1 - y_i) \log[1 - P(x_i)]\}$. Here we get

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

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0 \quad \text{and} \quad \sum_{i=1}^n x_{ij} [y_i - \pi(x_i)] = 0 \quad \text{for } j=1, 2, \dots, k. \text{ Let } \hat{\beta} \text{ denote}$$

the solution to these equations. Thus, the fitted values of multiple logistic

regression model are $\hat{P}(x_i)$. Then, the method of estimating the variance and co-variances of these estimated coefficients follows from the well developed theory of maximum likelihood estimation (6-7). This theory states that estimators are obtained from the matrix of partial derivatives of the log likelihood function. Further, the estimated standard errors of the estimated coefficients of

logistic regression model is given by $SE(\hat{\beta}_j) = \sqrt{\hat{V}ar(\hat{\beta}_j)}$ for $j = 0, 1, 2, \dots, p$. Alternatively, the Wald or Z statistic i.e. $W_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$, $j = 0, 1, 2, \dots, p$

is commonly used to test the significance of individual logistic regression coefficients for each independent variable. The multivariable analog of the

Wald test is $N = \hat{\beta}'_j [\hat{V}ar(\hat{\beta})]^{-1} \hat{\beta}_j = \hat{\beta}' (X'VX) \hat{\beta}$. The statistic N is

distributed as chi-square with $(k+1)$ degrees of freedom under the hypothesis

that each of the $k+1$ coefficient is equal to zero. The multivariable analog of the score for the significance of the model is based on the distribution of p derivatives of $L(\beta)$ with respect to β .

It is distributed as chi-square with 1 degree of freedom. The Wald statistic is simply the square of the (asymptotic) t -statistic. The Wald statistic can be used to calculate a confidence interval for β_j . We can assert with $100(1-\alpha)\%$ confidence intervals that the true parameter lies in the interval with boundaries $\hat{\beta} \pm Z_{\alpha/2} \text{ASE}$, where ASE is the asymptotic standard error of logistic $\hat{\beta}$. Parameter estimates are obtained using the principle of maximum likelihood; therefore hypothesis tests are based on comparisons of likelihoods or the deviances of nested models. The likelihood ratio test uses the ratio of the maximized value of the likelihood function for the full model (L_1) over the maximized value of the likelihood function for the simpler model (L_0). The likelihood-

ratio test statistic equals:
$$-2 \log \left(\frac{L_0}{L_1} \right) = -2[\log(L_0 - L_1)] = -2(L_0 - L_1).$$

This log transformation of the likelihood functions yields a chi-squared statistic. This is the recommended test statistic to use when building a model through backward elimination procedure. The purpose of logistic regression is to correctly predict the category of outcome for individual cases using the most parsimonious model. To accomplish this purpose, a model is created that includes all predictor factors that are useful in predicting the response variable.

Analysis and Interpretation

The dichotomized dental caries and periodontal disease data are analyzed. We have computed measures of central tendency and dispersion for both continuous and nominal variables. The multiple logistic regression model is constructed between the binary response variables i.e. dental caries (DMFT) and periodontal disease (CPITN) with factors independently.

The model estimation, in the first step, the multiple logistic full model is constructed for considering all factors and in the second step, the stepwise called multiple logistic reduced model is performed by considering only significant factors 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 probability of additional information gained from including further variables was less than 0.05. Lastly, 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, 95% confidence intervals and p-values of each factor computed.

Results

Comparisons in terms of parameter estimates, log likelihood and AIC values in particular are carried, discussed and presented. A total of 1760 subjects are included in the study with mean age is 34.26 ± 9.25 .

Model for dental caries (DMFT Index) and Periodontal Disease CPITN Index

Out of 21 factors, 14 factors are found to be statistically significant factors of dental caries ($p < 0.05$) such as age (in years), frequency of sweet consumption, mouth rinsing habit, smoking habit and alcohol habit, caste, socio-economic status, family size, sources of drinking water, time for sweet consumption, oral hygiene habits, frequency of brushing, timings of cleaning teeth and type of toothpastes. However, out of 21 factors, only five factors are found to be statistically significant factors of periodontal disease ($p < 0.05$) such as gender, frequency of brushing, timings of cleaning the teeth and type of toothpastes and family size. These significant factors have different signs of influence (Table 1).

Further, a log likelihood value for DMFT index model is -1013.1364 and CPITN index model is -1085.7876 . The Akaike's Information Criterion (AIC) value is 1.1752 in DMFT index model and 1.2577 in CPITN index model. But, main goal is to estimate the best fitting model for dental caries and periodontal disease, while minimizing the number of factors. The reduced regression model is obtained by excluding the factors which are not significant in full model. But respectively only 13 and 5 factors are found to be significant predictors of dental caries and periodontal disease in reduced models (Table 1). The log likelihood and AIC value of the reduced regression model are -1019.8106 , 1.1748 (DMFT index) and -1098.4320 and 1.2539 (CPITN index) respectively. Based on log likelihood and AIC values, the reduced logistic regression model is better fits for DMFT index data and the full and reduced logistic regression models are have approximately similar fit for CPITN model.

Odds of Risk Factors for Dental Caries

The estimated odds ratio of age (OR=1.1532, 95% CI: 1.0642-1.2495), caste (OR=0.8464, 95% CI: 0.7414-0.9662), socio-economic status (OR=0.7841, 95% CI: 0.6610-0.9302), family size (OR=0.6953, 95% CI: 0.5034-0.9603), sources of drinking water (OR=0.1427, 95% CI: 0.0945-0.2156), time for sweet consumption (OR=0.4444, 95% CI: 0.2130-0.9270), frequency of sweet consumption (OR=4.0962, 95% CI: 1.8070-9.2857), oral hygiene habits (OR=0.7522, 95% CI: 0.6068-0.9323), frequency of brushing (OR=0.5273, 95% CI: 0.3864-0.7195), timings of cleaning the teeth (OR=0.5809, 95% CI: 0.4362-0.7735), type of toothpastes (OR=0.5358, 95% CI: 0.4030-0.7123), mouth rinsing habit (OR=1.7898, 95% CI: 1.4137-2.2659), smoking habit (OR=1.6261, 95% CI: 1.2379-2.1362) and alcohol habit (OR=3.8206, 95% CI: 2.4047-6.0) found to have significant association with dental caries ($p < 0.05$). It means that, the respondents have higher risk of dental caries if they belong to older age, backward caste, low SES, lived in a family with >5 family members, drinking tube well and hand pump water, having sweet during and between meals, brushing their teeth with their fingers in morning or night only without fluoridated toothpastes with smoking habit and alcohol habit compared to their counterparts. The strength of association of some covariates with dental caries has significantly improved in reduced model.

Odds of Risk Factors for Periodontal Disease

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 influence 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 in a morning and night without pastes/powder have higher prevalence of periodontal disease as compared to their counterparts. However, there is an improvement in the strength of association among some covariates seen in reduced model.

Table 1: The Estimated Coefficients of Factors from Full and reduced Logistic Regression Model to Dental Caries and Periodontal Disease Dichotomous Data

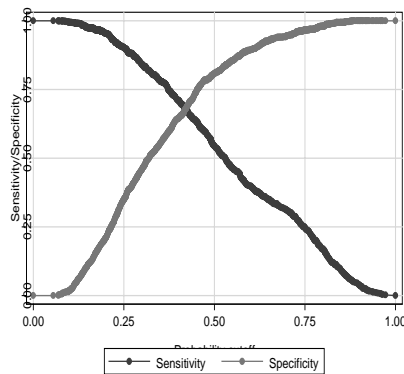
Factors	Dental Caries				Periodontal disease			
	Full model		Reduced model		Full model		Reduced model	
	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
Constant	4.57	0.89*	1.17	0.05*	1.71	0.84	1.14	0.22*
Gender	0.01	0.12	-	-	0.41	0.12*	0.40	0.10*
Age (in years)	0.14	0.04*	0.85	0.06*	0.05	0.04	-	-
Religion	0.18	0.12	-	-	-0.02	0.11	-	-
Caste	-0.17	0.07*	0.77	0.06*	-0.06	0.06	-	-
Socio-economic status	-0.24	0.09*	0.71	0.12*	-0.12	0.08	-	-
Family size	-0.36	0.16*	0.18	0.03*	0.30	0.16*	0.29	0.16*
Staple food	-0.20	0.19	-	-	0.05	0.18	-	-
Sources of drinking water	-1.95	0.21*	2.05	0.50*	-0.30	0.20	-	-
Dietary habits	0.12	0.11	-	-	-0.15	0.11	-	-
Time for sweet consumption	-0.81	0.38*	1.75	0.08*	0.10	0.32	-	-
Frequency of sweet consumption	1.41	0.42*	0.55	0.09*	0.36	0.37	-	-
Oral hygiene habits	-0.28	0.11*	0.59	0.08*	-0.07	0.10	-	-
Frequency of brushing	-0.64	0.16*	0.59	0.07*	-0.35	0.14*	-0.28	0.14*
Timings of cleaning the teeth	-0.54	0.15*	1.77	0.20*	-0.31	0.14	-0.13	0.13*
Methods of brushing	-0.22	0.12	-	-	-0.02	0.12	-	-
Materials used for brushing their teeth	-0.25	0.18	-	-	0.17	0.17	-	-
Type of toothpastes	-0.62	0.15*	1.51	0.20*	-0.47	0.14*	-0.39	0.11*
Mouth rinsing habit	0.58	0.12*	3.02	0.35	0.01	0.12	-	-
Smoking habit	0.49	0.14*	1.11	0.12	0.02	0.13	-	-
Chewing habit	0.27	0.23	-	-	0.22	0.21	-	-
Alcohol habit	1.34	0.24	-	-	0.32	0.21	-	-
Log likelihood	-1013.14	-1019.81			-1085.79	-1098.43		
AIC	1.18	1.17			1.26	1.25		

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

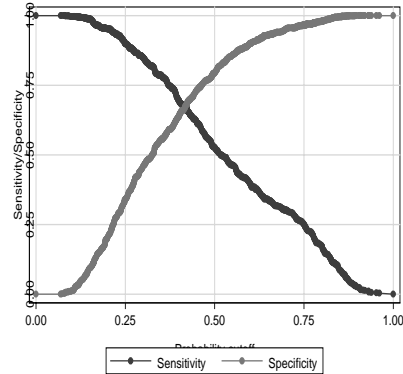
The plot of sensitivity and specificity versus criterion value and accuracy of the test in terms of ROC for the response variables (DMFT Index and CPITN Index) in the full and reduced model is presented in the following Figure 1.

Figure 1: The plot of sensitivity and specificity versus criterion value for the response variables (DMFT Index and CPITN Index) in the full and reduced models

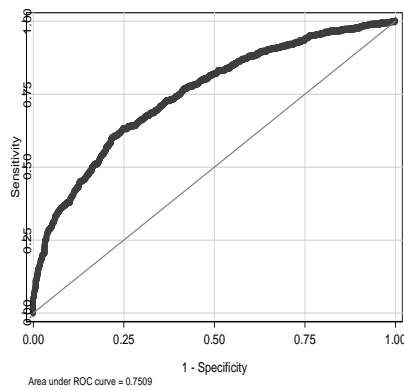
Full model for DMFT Index



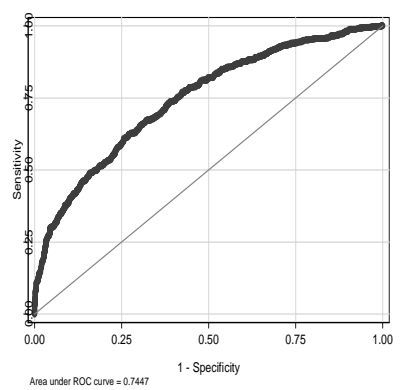
Reduced model for DMFT Index



Full model for CPITN Index



Reduced model for CPITN Index



The area under ROC curve of the response variable i.e. DMFT Index in the full model is 0.7509 and in the reduced model is 0.7447. It provides a summary of the accuracy of the diagnostic test which is approximately 75% in full model and 74% in reduced model.

Similarly, the area under ROC curve of the response variable i.e. CPITN Index in the full model is 0.7509 and for the reduced model is 0.5821. It provides a summary of the accuracy of the diagnostic test which is approximately 61% in full model and 58% in reduced model.

Discussion and Conclusions

Different statistical methods i.e. regression analysis, multilevel modelling, logistic regression and ordinal regression techniques employed to analyze the dental caries data (DMFT Index) and periodontal disease (CPITN Index) data yield results having different focuses. The regression methods allow researchers to identify factors related to social-economic-demographic and other factors related to oral health that contribute to overall status of response variables. These methods also permit researcher to estimate the magnitude of the effect of factors. Usually, the nature of dental caries data is in count scale and periodontal disease data is in ordinal scale. Therefore, the dental caries and periodontal disease data are converted into binary or dichotomous outcomes (with and without disease). Use of a dichotomous outcomes in traditional multiple linear regression model violates the assumption of normality and homoscedasticity. Hence, the main aim of the present chapter is to utilize the applications of logistic regression model to assess the effect and relationship between factors and binary response variables independently. Impact of different factors (determinants) on a dental caries has already been investigated in several studies. The strength of the present study is the inclusion of a large number of detailed probable and prominent factors or determinants for modelling dental caries.

The findings shows that there are strong positive associations existed between age (in years), frequency of sweet consumption, materials used for brushing the teeth and smoking habit with dental caries. These findings coincide with a study done by Eija and Leena (11) except for the significant association between total sugar consumption and high risk of occurrence of dental caries only when oral hygiene is poor. The socio-economic status, oral hygiene habits, frequency of brushing, type of toothpastes, duration of change of toothbrush are significant factors having negative association with dental caries.

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., (12) and Grossi et al., (13-14). Males usually exhibit proper oral hygiene than females (15). 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 (12-14, 16-18). Also it is found that the severity of the disease is more with respect to plaque development and gingivitis in elderly persons as compared to younger persons (18).

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 (19). However, in this study an association is not found to be statistically significant. But, the Ramfjord et al. (20) 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 (21-23) and prevalence of Acute Ulcerative Gingivitis (ANUG) was demonstrated as early in 1946 (23, 24-25). 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 (17, 23, 25-33). 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 chapter, we compared performance of full logistic model with that of reduced logistic model using log likelihood estimate for both DMFT and CPITN index data. The results show that reduced logistic regression model is slightly a better fit as compared to full logistic regression model applied to dichotomized DMFT index data. However, the fitting performance of full logistic regression model is slightly better as compared to reduced logistic regression model applied to dichotomized CPITN index data.

References

- World Health Organization (WHO). Oral Health Surveys. Basic Methods. WHO Geneva, 1997.
- Ainamo J., Barnes D., Beagrie G. et al. Development of the World Health Organization (WHO) community periodontal index of treatment needs (CPITN). *Int. Dent. J.*, 1982, 32(3); 281-291.
- Nelder, J. A. E. Wedderburn, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society A*, local, 1972, 135; 370-384.
- Agresti A. Tutorial on modeling ordered categorical response data. *Psychological Bulletin*, 1989, 105; 290-301.
- Cox, D. R. and Snell, E. J. *Analysis of Binary Data*, Second Edition, Chapman and Hall, London, 1989.
- Rao C. R. *Linear Statistical Inference and its application*, second edition, Wiley Inc., New York, 1973.
- David W. H. and Stanley. *Applied Logistic Regression*, Second edition, A Wiley Inter science Publication, John Wiley and Sons, Inc. New York, 2000.
- Metz C. E. Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 1978; 8, 283-298.
- Zweig M. H. and Campbell G. Receiver-operating characteristic (ROC) plots, a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 1993; 39; 561-577.
- Griner P. F., Mayewski R. J., Mushlin A. I. and Greenland P. Selection and interpretation of diagnostic tests and procedures. *Annals of Internal Medicine*, 1981, 94; 557-592.
- Eija K. K. and Leena R. Relationship of oral hygiene and sugar consumption to risk of caries in children. *Community Dent. Oral Epidemiol.*, 1982, 10; 224-233.
- Miller A. J., Brunelle J. A., Carlos J. P., Brown L. J. and Loe H. Oral Health of United States Adults. National findings. Bethesda MD. National Institute of Dental Research. NIH publication No. 87, 1987, 2868.
- Grossi S. G., Zambon J. J., Ho A. W. et al. Assessment of risk for periodontal disease-I, Risk indicators for attachment loss. *J. Periodontology*, 1994, 65; 260-267.
- Grossi S. G., Genco R. J., Machatei E. E. et al. Assessment of risk for periodontal disease. II. Risk indicators for alveolar bone loss. *J. Periodontology*, 1995, 66; 23-29.
- U. S. Public Health Service, National center for Health Statistics. 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, 1979.
- Marshal-Day C. D., Stephens R. G. and Quigley L. F. Jr. Periodontal disease prevalence and incidence. *J. Periodontology*, 1955, 26; 185-203.
- Schei O., Waerhaug J., Lovdal A. and Arno A. Alveolar bone loss as related to oral hygiene and age. *J. Periodontology*, 1959, 30; 7-16.
- Abdellatif H. M. and Burt B.A. An epidemiological investigation into the relative importance of age and oral hygiene status as determinants of periodontitis. *J. Dent. Res.*, 1987, 66; 13-18.
- Russell A. L. *Geographical Distribution and Epidemiology of Periodontal Disease*. Geneva. World Health Organization, (WHO/DH/33/34), 1960.
- Ramfjord S. P., Emslie R. D. and Waerhaug J. Epidemiological studies of periodontal disease. *Am. J. Public Health*, 1968, 28; 1713-1722.
- Pindborg J. J. 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.*, 1947, 26; 261-264.

-
- Frandsen A. and Pindborg J. J. Tobacco and gingivitis III. Difference in action of cigarette and pipe smoking. *J. Dent. Res. Dent.*, 1949, 28; 464-465.
 - Solomon H. A., Priore R. L. and Bross I. D. Cigarette smoking and periodontal disease. *J. Am. Dent. Assoc.*, 1968, 77; 1081-1084.
 - Pindborg J. J. Tobacco and Gingivitis II. Correlation between consumption of tobacco, ulcer membranous gingivitis and calculus. *J. Dent. Res.*, 1949, 28; 460-463.
 - Bergstrom J. and Floderus-Myred B. Co-twin control study of the relationship between smoking and some periodontal disease factors. *Community Dent. Oral Epidemiol.*, 1983, 11; 113-116.
 - Preber H., Kant T. and Bergstrom J. Cigarette smoking, oral hygiene and periodontal health in Swedish army conscripts. *J. Clin. Periodontology*, 1980, 7; 106-113.
 - Bergstrom J. Short term investigation on the influence of cigarette smoking upon plaque accumulation. *Scand. J. Dent. Res.*, 1981, 89; 235-238.
 - Bergstrom J. and Floderus-Myred B. Co-twin control study of the relationship between smoking and some periodontal disease factors. *Community Dent. Oral Epidemiol.*, 1983, 11; 113-116.
 - Bergstrom J. Oral hygiene compliance and gingivitis expression in cigarette smokers. *Scand. J. Dent. Res.*, 1990, 98; 497-503.
 - Sheiham A. Periodontal disease and oral cleanliness in tobacco smokers. *J. Periodontology*, 1971, 42; 259-263.
 - Macgregor I. D. M., Edgar W. M. and Greenhood A. R. Effects of cigarette smoking on the rate of plaque formation. *J. Clin. Periodontology*, 1985, 12; 35-41.
 - Preber H. and Bergstrom J. Cigarette smoking in patients referred for periodontal treatment. *Scand. J. Dent. Res.*, 1986, 94; 102-108.
33. Herulf G. On the marginal alveolar ridge in adults. *Sven Tandlak Tidsk*, 1968, 63; 675-703.