Challenges in dental statistics: data and modelling

Domenica Matranga(1), Paolo Castiglia(2), Giuliana Solinas(2)

ABSTRACT

The aim of this work is to present the reflections and proposals derived from the first Workshop of the SISMEC STATDENT working group on statistical methods and applications in dentistry, held in Ancona (Italy) on 28th September 2011. STATDENT began as a forum of comparison and discussion for statisticians working in the field of dental research in order to suggest new and improve existing biostatistical and clinical epidemiological methods.

During the meeting, we dealt with very important topics of statistical methodology for the analysis of dental data, covering the analysis of hierarchically structured and over-dispersed data, the issue of calibration and reproducibility, as well as some problems related to survey methodology, such as the design and construction of unbiased statistical indicators and of well conducted clinical trials.

This paper gathers some of the methodological topics discussed during the meeting, concerning multilevel and zero-inflated models for the analysis of caries data and methods for the training and calibration of raters in dental epidemiology.

Key words: Dmft; Multilevel modeling; Zero-inflation; Overdispersion; Reproducibility; Examiner reliability

(1) Department of Sciences for Health Promotion and Mother and Child Care, University of Palermo, Palermo, Italy
(2) Department of Biomedical Science, University of Sassari, Sassari, Italy

CORRESPONDING AUTHOR: *Domenica Matranga, University of Palermo, Via del Vespro, 133, 90127 Palermo, Italy. Tel: +39 091 6553604. Fax: +39 091 6553647.*
e-mail: domenica.matranga@unipa.it
DOI: ***10.2427/8759***

HIERARCHICAL STRUCTURED DATA IN DENTAL DATA

The current etiologic models of dental caries based on dynamic processes taking place just within the individual oral cavity may partly explain the caries etiology.

With regards to oral health, individual level factors have not fully explained the prevalence of an array of diseases important to public health. For example, dental caries, which is the most prevalent chronic disease, has spread worldwide affecting people of all ages.

The biological causation alone does not provide sufficient explanation for the variations in the experience of caries among different populations [1]. Determinants of individual health are not always the determinants of population health [2], and dental researchers have not yet linked macro social forces, such as neighborhood characteristics, with patterns of oral health status and disease in populations. Therefore, researchers' interest in detecting dental caries is reawakening from the point of view of neighborhood characteristics [3, 4] in order to shape individually based risk factors and behaviours [5-8]. Few published data concerning social differences in dental health have been reported in Italy to date [9].

Dental data are a typical hierarchical situation due to the clustered nature of disease, within surfaces, teeth, individuals, as well as group clusters such as schools, dental practices and geographical areas.

An example of a multilevel modeling approach to dental caries

During the year 2003, cross-sectional data were collected by the Italian Pathfinder survey. The purpose of this study was to describe the oral health status of a population of 12-year-old children using the DMFT index (Decayed, Missing, and Filled Teeth), as indicator of dental health status proposed by the World Health Organization [10].

Despite the fact that the DMFT resulted close to the global goal set (DMFT≤1), unequal geographical distribution of DMFT among Italian regions observed (North-Western, North Eastern, Central, Southern and Insular) suggests that it is worth combining the role of individual and community factors in order to investigate their contribution to the DMFT variability in people living in different Italian regions.

Multilevel regression analysis including 3 199 children was applied involving two levels of analysis and assuming that individual observation (Level-1) are nested within Italian areas (Level-2). With regards to individual characteristics, gender, parents' educational level, frequency of tooth-brushing habits, intake of sweet foods, sweet drinks before going to sleep, soft drinks and supply of fluoride were proposed as predictors of the outcome of the DMFT index in the multilevel analysis. These variables were combined with community data obtained by ISTAT in five Italian regions [11], including average income (in euros), area of dwelling per person (m^2), unemployment rate (%) and number of dentists (per 100 000). During dental check-ups carried out by trained and calibrated examiners, data on oral health status were collected.

A sequence of three hierarchical linear models was explored using GLAMM running in STATA 9 [12].

First, an unconditional model was used to explain if there was significant variation in the caries severity measure between five Italian regions.

The equation, called null model:

$$DMFT_{ij} = \beta_0+(\mu_j+e_{ij})$$

contains no predictors, and the DMFT of a child depends only on the mean level of all children in all Italian areas as (β_0) and on a differential for each area (μ_j) and for each child (e_{ij}). The DMFT is the continuous outcome variable for the *j*th person in the *i*th area. This model allowed partitioning of the total variance of the outcome within group variance (individual level) and between group variance (area level). The results indicate that there is a large variability in DMFT at an individual level (σ^2=2.84, $p<0.001$) and far less at the area level (σ^2=0.04, $p<0.001$). The total variation between regions (only 1.4%) was much smaller than the variability at an individual level (98.6%), nevertheless it was statistically significant ($p<0.01$).

Secondly, individual level predictors were included to model the outcome in a variance component model. This model helped to explain the proportion of caries variance within five Italian regions which were accounted for, once the individual predictors were examined.

The inclusion of these predictors contributed only 5.98% of the individual level variance in DMFT. Gender inequality on DMFT indicated that girls have a significantly higher mean DMFT value than boys (b=0.20, p=0.03). Furthermore, children who eat sweet foods frequently and drink milk and sugary beverages before going to sleep had a significantly higher mean DMFT (b=0.09, p=0.003) b=0.06, p=0.03), respectively. Also, toothbrushing habits ≥2 times a day (b=-0.21, p=0.001) and supply of fluoride (b=-0.15, p=0.015) were significantly associated with a decrease in DMFT. Higher education levels of parents (b=-0.28, p<0.001) contributed to DMFT decrease.

Finally, in the third model:

$$DMFT_{ij} = \beta_0 + \beta X_{ij} + (\mu_j + e_{ij})$$

community predictors were added to the second model to assess whether they had any influence in explaining the variability of caries among Italian regions. The parameters β assess the relationship between dental health, the individual and community variables across all area sections.

In the final model, the unexplained variance reduction of DMFT between areas was estimated. It was found that 97.3% of DMFT variance by section-areas was explained by community variables. Only the area of dwelling per person significantly contributed to DMFT decrease (b=0.09, p=0.001).

The use of multilevel analysis in dental data is an interesting statistical approach for cluster data. For this reason, this analysis was proposed for the Italian survey study on DMFT index to correct the bias in the variance estimates.

The results obtained by multilevel analysis highlights that as well as individual characteristics, the community context may also influence DMFT index. This suggests that a deeper investigation is required on gender differences in dental health, which may imply social features with equitable consequences.

ZERO-INFLATED MODELS FOR CARIES DATA

In dental caries research, a well established indicator of individual oral health condition is the dmft (DMFT) index, obtained as the total number of decayed (d/D), missing (m/M) and filled (f/F) deciduous milk teeth (t) or permanent teeth (T). In developed countries, great attention is given to oral hygiene and oral health care, so an excess number of zero dmft/DMFT counts is likely. There is a dual data generating process which can be modeled as a mixture of two distributions: a zero-distribution from which only zero values are observed and a Poisson distribution (Binomial in case of bounded data) from which all of the nonzero and a few of the zero values are observed. Therefore, there are two sources of zero counts: a few coming from the first distribution are *structural zeros*, others coming from the count distribution are *sampling zeros*. Some children may never experience caries (structural zeros), while others may have recorded zero dmft in the sample time frame, and a non-zero in some future time frame (sampling zeros). The overall probability of zero counts is the combined probability of zeros from two groups; however which group they come from is not known.

The existence of this dual process causes an unobserved population heterogeneity, such that a single Poisson (binomial) parameter μ is not sufficient to describe the population.

One approach to solve this problem is to assume that the heterogeneity involved in the data can be adequately described by modeling μ as a random variable. If the process is Poisson, the population mean μ follows a gamma distribution and the mixture density is the negative binomial distribution. If the process is binomial (bounded counts), the success probability μ follows a beta distribution and the mixture density is the beta-binomial distribution.

Another approach to solve population heterogeneity is based on zero-inflation, that is to model data as mixture of the count distribution, with a non-zero central location, and a degenerate distribution at zero, with a zero central location. The non-zero central location and the mixing proportion parameter are of interest.

The two approaches can be combined to model both over-dispersion and excess zero, giving rise to zero-inflated negative binomial and zero-inflated beta-binomial models.

The seminal paper of statistical methods for over-dispersed data in dental caries epidemiology was written by Böhning et al. in 1999 [13]. They proposed Zero-inflated Poisson (ZIP) regression models to analyze the BELCAP (Belo Horizonte Caries Prevention) study data-set of 797 Brazilian seven year-old schoolchildren. As highlighted by the authors themselves and deepened by Skrondal and Rabe-Hesketh [14] and by Giltorpe et al. [15], zero-inflated Binomial (ZIB) regression models could be more appropriate because the dmft/DMFT index is bounded between zero and the maximum number of teeth in the mouth (20 deciduous and 32 permanent). In particular, the study of Giltorpe et al. [15] proposed an association of model selection and interpretation to a priori hypothesis of data generation, by distinguishing different scenarios regarding latent risks of disease onset and disease progression.

Karlis and Ntzoufras [16] proposed a Zero-inflated Poisson Difference (ZDP) model for correlated paired count data and showed an interesting application of their method to the difference between the dmft index before and after treatment of the BELCAP data set. The ZDP could model the difference of paired data and capture an excess of zero-values. The authors gave a clinical interpretation of zero differences, which could be resumed in Figure 1.

Preventive treatment is effective because unchanged dmft means that a child's caries do not worsen (sampling zero) or that a zero dmft is still zero after treatment (structural zero). On the other hand, therapeutic treatment is ineffective because unchanged dmft means that either a child's caries do not improve (sampling zero) or that a chronic non-zero dmft is stationary after treatment (structural zero).

An open line of research attains the development of zero-inflated multilevel modeling. Solinas et al. [17] proposed a multilevel ZINB model, while Burnside [18] proposed a multilevel ZIB model for binary response estimated using LatentGold software. As indicated by the author himself, one limitation of this software was that the maximum that could be taken into account is a two level hierarchy.

An example of modeling over-dispersed dental caries data

A cross-sectional sample of 511 schoolchildren of the city of Palermo (Sicily Region, Italy), 153 of whom were aged 5 (29.94%) and 358 of whom were aged 12 (70.06%), was selected following the cluster sampling technique from a stratified population [19]. The aim of the study was to assess the prevalence of caries and the relation between their dmft index and their oral health behaviours and socio-economic factors. In the published paper, the continuous DMFT/dmft index was categorized as 0 (absence of caries) and 1 (presence of caries) and a logistic regression model was estimated. Alternatively, we show the results of modeling both over-dispersion and excess zero in 5 year-old children.

The distribution of dmft/DMFT was non-normal, highly skewed (Skewness=2.47) and contained excessive zeros compared with standard Poisson distribution (% of observed and predicted zeros, respectively, 0.53 and 0.18). The mean of dmft/DMFT index was 1.73 (SD=2.83).

At the univariate analysis, the dmft/DMFT index was significantly associated with gender ($p=0.042$), mother's educational level ($p=0.002$), mother's employment status ($p=0.002$) and dental visits ($p=0.042$). Only dental visit attendance and mother's educational status were confirmed statistically significant at the multivariate analysis, with a decreased risk of caries for mid-level against low-level educated mothers. The increased risk for children attending dental visits can be explained as these visits took place for therapeutic reasons (Table 1).

On the basis of the LR test, the over-dispersed NB was more appropriate than the Poisson regression model ($p<0.001$). The Vuong test to compare the zero-inflated negative binomial (ZINB) model against the NB model was not statistically significant ($p=0.501$). For this reason modelling excess zeros was not considered in this application.

Our results are in agreement with our previous published study [19] where logistic model for dmft in preschool children showed that mid-level education of the mother was a strong protective factor in caries prevention in children (OR=0.07, 95% CI: 0.02–0.35). In addition, modelling over-dispersion and excess zero of dmft could reveal other significant explanatory variables, associated with behavioural and social factors.

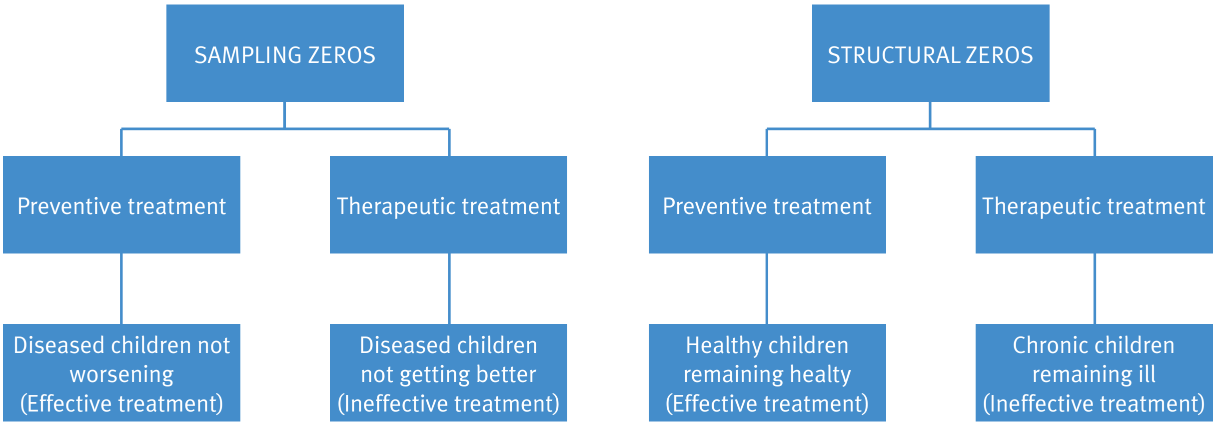
TRAINING AND CALIBRATION OF RATERS FOR EPIDEMIOLOGICAL STUDIES IN DENTISTRY

An important aspect of any survey is the use of appropriate methodologies to reduce the effects of potential confounding factors.

In the context of community dentistry, where multicenter studies are often conducted, a matter of great concern is the variation in disease diagnosis between two or more examiners. Therefore, multiple examiner investigation has long been considered an essential issue in this type of epidemiological study. For this reason, training is a necessary step in study design and was defined

FIGURE 1

INTERPRETATION OF ZERO DIFFERENCES $Z = DMFT_{after} - DMFT_{before}$



by the *Guidance on the Statistical Aspects of Training and Calibration of Examiners for Surveys of Child Dental Health by British Association for the Study of Community Dentistry* (BASCD), which teaches the agreed interpretation of diagnostic criteria [20].

In addition to training, that implies a correct presentation, interpretation, explanation and discussion of clinical examples in a timely manner and in a sympathetic atmosphere; and calibration which allows a formal measure of how well the examiner can interpret the criteria in comparison to the *gold standard*, must be performed [10].

Since calibration and training become extremely important if more than one examiner is involved in data collection, it is necessary to ensure that not only each examiner is consistent and reliable as an individual (intra-examiner reliability) but especially that each one agrees with the others (inter-examiner reliability). As reliability is related to the extent at which examiners agree in their evaluations, it is consequently investigated by examining several patients twice, to discover if diagnostic results for a rater remain reliable in both measurements. Intra-examiner reliability is calculated on measurements done by each examiner on the same patients and it might be affected by the presence of some systematic biases, mainly due to the difficulty of blinding the rater in order to guarantee the independence of the two measurements. Therefore, inter-examiner evaluation might be a more accurately estimated reliability, as already described in other studies of this kind [21].

Some international guidelines recommend several statistical methods in order to evaluate reliability both for dichotomous and quantitative variables [10, 20, 22], each of them showing specific characteristics but also their respective shortcomings.

For dichotomous variables, Dice's coefficient (or Sørensen similarity index) and "percent agreement" have largely been used, as the proportion of agreement between examiners judging (e.g. the presence or absence of caries disease). In spite of their intuitive simplicity and directness, this kind of *agreement* overestimates reliability because it contains both *chance agreement,* which could occur if the rater has just guessed each rating, and *beyond chance agreement*. Therefore, Sensitivity and Specificity have alternatively been suggested. These methods are considered to represent the intrinsic capability of raters in scoring, referring to the gold standard, and thus to be independent of prevalence; however, this may not be the case in practice. Yet, these indexes might be influenced by changing prevalence of disease, i.e. the lower the binomial variability because of declining prevalence, the higher the probability of reaching a by chance agreement. It has been described as quite easy to reach an agreement among raters for low levels of caries disease [23] and that it might also be referred to an informative cluster size [24].

In order to correct for chance agreement, Cohen's kappa coefficient is used instead of percent agreement. The range of possible values of kappa is from -1 to 1, even if usually values between 0 (chance agreement) and 1 (perfect agreement) are employed. A negative kappa would indicate less agreement than that expected by chance. Nevertheless, kappa statistics also includes disadvantages, like the above mentioned influence of disease distribution and the characteristics of unit of analysis (subject,

TABLE 1

NEGATIVE BINOMIAL MODEL FOR THE DMFT IN 5 YS CHILDREN			
DMFT COMPONENT	IRR	95%CI	P
GENDER FEMALE VS MALE	1.32	[0.73; 2.40]	0.354
MOTHER'S EDUCATIONAL LEVEL MIDDLE VS LOW HIGH VS LOW	0.33 0.31	[0.14; 0.76] [0.06; 1.64]	0.009 0.168
MOTHER'S EMPLOYMENT STATUS HOUSEWIFE VS EMPLOYEE UNEMPLOYED VS EMPLOYEE	0.63 1.20	[0.27; 1.47] [0.48; 2.98]	0.285 0.699
DENTAL VISIT ATTENDANCE YES VS NO	4.52	[1.95;10.50]	<0.001
OVER-DISPERSION	1.92	[1.26;2.94]	
VUONG TEST OF ZINB VS NB			0.501

tooth, surface level) [25]. Moreover, it cannot be applied to non-categorical data. In this case, Pearson's correlation should be used. However, if a systematic bias is present (e.g. if a rater indicates a lesion on one or more teeth per patient that the other rater does not report) Pearson's correlation coefficient overestimates the agreement [26]. In dentistry, this phenomenon may occur, for example in the presence of previous campaigns of sealing teeth, in case a rater systematically misreads the sealant as a white filling.

Therefore, the statistic might be biased and, moreover, as in dentistry the number of considered units is usually higher than in other kind of clinical studies (teeth or surfaces instead of subjects), even the confidence interval of Pearson's statistic, usually narrow because of the great number of studied units, might be uninformative. For quantitative variables other statistical methods have been suggested. Traditionally, an analysis of variance for fixed effect can be applied [27]. However, it might fail to enhance significant differences among raters in case of high probability of chance agreement [23]. If quantitative indexes are used, like DMFT/S-dmft/s, the intraclass correlation, the concordance correlation coefficient or the Bland and Altman method can be used. Otherwise, if an assumption that continuous latent variables underlying the contingency table exist, a tetrachoric or polychoric correlation should be used [28].

Other methods have been suggested and new advanced statistical modelling appear promising in the near future.

In conclusion, some important points must be checked in preparing a caries experience survey: a) training should be carried out following international guidelines depending on the characteristics of the study; b) standardized methods must be used; c) a gold standard should be used and the practitioner should be subjected to regular quality control; d) reliability measures should be implemented according to the characteristics of the study variables and population; e) different methods should be used to evaluate reliability; f) the classifications for observed statistics (strength of kappa or correlation coefficients) should be well declared; g) confidence intervals should always be presented.

For all these reasons, a reproducibility study must precede a multicenter epidemiological study in order to evaluate the inter-examiner reliability, and these finding should be published alongside the survey results.

Unfortunately, there are no absolute guidelines available to date and the researcher should take all the above mentioned items into proper consideration, comparing them with the most recent literature.

CONCLUSIONS

The goal for this workshop was to improve our understanding of knowledge of statistical methods and applications in dentistry and how statistical techniques could be used in dental research.

The contributions of the invited speakers at the workshop have highlighted the complexity of dental data. The hierarchical structure of such data is the reason it is inappropriate to apply classical regression methods when the objective of research is to examine the relationship between an outcome and (more than) one covariate (s). Alternatively, to take into account the correlated nature of clustered dental data the multilevel analysis can be used. In addition, the zero-inflated models have proved to be attractive methodological approaches to study the excess number of zero dmft/ DMFT index.

The accuracy of diagnostic tests on different aspects of collection of data, as well as good agreement between different observers were revealed to be important processes in data quality control, to ensure reliable oral health research.

We hope that the success of the Workshop will be the beginning of new collaborations between dental researchers and biostatisticians to create a multidisciplinary network to improve the methodological quality of dental research.

ACKNOWLEDGEMENTS: *we express our gratitude to Flavia Carle, the President of the VI National Congress of SISMEC, for permitting us to organize the Satellite Workshop "New Frontiers in Statistics for Dental Data". A special thanks goes to Girvan Burnside, Elisabetta Petracci and Guglielmo Campus for having collaborated to ensure a successful Workshop. Protecter & Gamble S.R.L contributed to the sponsorship of the Workshop.*

References

[1] Holst D. Caries in populations-a theoretical causal approach. Eur J Oral Sci. 2001; 109(3): 143-8
[2] Kawachi I, Berkman LF. Social Epidemiology. Oxford. New York; Oxford University Press 2000
[3] Link BG, Phelan J. Social conditions as fundamental causes of disease. J Health Soc Behav. 1995; Spec No: 80-94
[4] Pattussi MP, Hardy R, Sheiham A. The potential impact of neighborhood empowerment on dental caries among adolescents. Community Dent Oral Epidemiol 2006; 34(5): 344–50
[5] Campus G, Lumbau A, Lai S, et al. Socio-economic and behavioural factors related to caries in twelve-year-old Sardinian children. Caries Res. 2001; 35(6): 427-34
[6] Moynihan P. The scientific basis for diet, nutrition and prevention of dental diseases. Background paper for the Joint WHO/FAO Expert Consultation on diet, nutrition and the prevention of chronic diseases. Geneva; WHO 2002
[7] Rugg-Gunn A, Nunn JH. Nutrition, diet, and oral health. Oxford. New York; Oxford University Press 1999
[8] Sheiham A. Dietary effects on dental diseases. Public Health Nutr 2001; 4: 569–91
[9] Angelillo IF, Nobile CGA, Pavia M. Oral health status and treatment needs in immigrants and refugees in Italy. Eur J Epidemiol 1996; 12: 359-65
[10] WHO (World Health Organization). Oral Health Surveys: Basic Methods. Geneva; WHO 1997
[11] ISTAT. Codici dei Comuni delle Province e delle Regioni. http://www.istat.it/it/archivio/6789. Accessed November 2011
[12] Rabe-Hesketh S, Skrondal A. Multilevel and Longitudinal Modeling using Stata. College Station, TX; Stata Press 2005
[13] Böhning D, Dietz E, Schlattmann P, et al. The Zero-Inflated Poisson Model and the Decayed, Missing nd Filled Teeth Index in Dental Epidemiology. JRSS Serie A 1999; 162(2): 195-209
[14] Skrondal A, Rabe-Hesketh S. Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models. London; Chapman & Hall 2004
[15] Giltorpe MS, Frydenberg M, Cheng Y, Baelum V. Modelling count data with excessive zeros: The need for class prediction in zero-inflated models and the issue of data generation in choosing between zero-inflated and generic mixture models for dental caries data. Stat Med 2009; 28: 3539-53
[16] Karlis D, Ntzoufras I. Bayesian analysis of the differences of count data. Stat Med 2006; 25: 1885-1905
[17] Solinas G, Campus G, Maida C et al. What statistical method should be used to evaluate risk factors associated with dmfs index? Evidence from the National

Pathfinder Survey of 4-year-old Italian Children. Community Dentistry and Oral Epidemiology 2009; 37(6): 539-546
[18] Burnside G, Pine CM, Williamson PR. The application of multilevel modelling to dental caries data. Statist. Med. 2007; 26: 4139–49
[19] Pizzo G, Piscopo MR, Matranga D, et al. Prevalence and socio-behavioral determinants of dental caries in Sicilian schoolchildren. Med Sci Monit, 2010; 16(10): PH83-89 24
[20] Pine CM, Pitts NB, Nugent ZJ. BASCD guidance on the statistical aspects of training and calibration of examiners for surveys of child dental health. Community Dent Health 1997; 14:18-29.28
[21] McPartland JM, Goodridge JP. Counterstrain and traditional osteopathic examination of the cervical spine compared. J Bodywork Movement Ther 1997; 1: 173-8
[22] Ismail AI, Sohn W, Tellez M, et al. The International Caries Detection and Assessment System (ICDAS): an integrated system for measuring dental caries. Community Dent Oral Epidemiol (2007); 35: 170–8
[23] Castiglia P, Campus G, Solinas G, et al. Children's Oral Health in Italy: Training and Clinical Calibration of Examiners for the National Pathfinder about Caries Disease. Oral Health Prev Dent 2007; 5(4): 255-61
[24] Williamson JM, Datta S, Satten GA. Marginal analyses of clustered data when cluster size is informative. Biometrics 2003; 59: 36–42
[25] Agbaje JO, Mutsvari T, Lesaffre E, Declerck D. Measurement, analysis and interpretation of examiner reliability in caries experience surveys:some methodological thoughts. Clin Oral Investig. 2012;16(1): 117-27
[26] Hunt RJ. Percent Agreement, Pearson's correlation, and Kappa as measures of inter-examiner reliability. J Dent Res 1986; 65: 128-30
[27] Fleiss JL. Reliability of measurements. In: Fleiss JL. The design and analysis of clinical experiments. New York; John Wiley and Sons 1986
[28] Uebersax JS. Statistical methods for rater agreement: the tetrachoric and polychoric correlation coefficient.2005. http://www.john-uebersax.com/stat/tetra.htm

*