Community Dent Oral Epidemiol 2009; 37: 539–546

COMMUNITY DENTISTRY AND ORAL EPIDEMIOLOGY

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

Giuliana Solinas[1], Guglielmo Campus[2], Carmelo Maida[2], Giovanni Sotgiu[1], Maria Grazia Cagetti[3], Emmanuel Lesaffre[4] and Paolo Castiglia[1]

[1]Institute of Hygiene and Preventive Medicine, Unit of Biostatistics, [2]Dental Institute University of Sassari, Sassari, Italy, [3]WHO Collaborating Centre of Milan for Epidemiology and Community Dentistry, University of Milan, S. Paolo Hospital, Milan, Italy, [4]Department of Biostatistics, Erasmus MC, Rotterdam, The Netherlands and Biostatistical Centre, Katholieke Universiteit, Leuven, Belgium

Solinas G, Campus G, Maida C, Sotgiu G, Cagetti MG, Lesaffre E, Castiglia P. 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 Dent Oral Epidemiol 2009; 37: 539–546.

Abstract – *Background:* Traditional approaches to the analysis of dmfs/DMFS count data pose analytical challenges, considering the increasing proportion of zeroes in the distribution. The aim of this paper was to predict the probability of 'caries-free' subjects and the dependence of dmfs index on the influence of childhood sociodemographic factors, through the application of regression models. *Methods:* Data were gathered as part of the National Pathfinder Survey of 4-year-old Italian children. Clinical data on caries disease (dmfs) and childhood sociodemographic factors were collected. The predicted probability for Poisson, negative binomial and zero-inflated models (Poisson and negative binomial) were estimated using STATA commands for count outcomes. The outcome variable in the regression models was the severity of the disease (dmfs index), while statistically significant variables on bivariate analysis were considered as covariates. *Results:* Out of 5538 children, 4344 (78.44%) had a dmfs = 0. The mean dmfs index was 1.36 (range: 0–104). The statistical significance of the dispersion parameter (O = 141.6, $P < 0.0001$) showed the inappropriateness of the Poisson model when compared with the negative binomial model. Vuong's test indicated that the zero-inflated models (ZIP and ZINB) fitted the data significantly better than the others ($P < 0.001$). A significative likelihood ratio statistic indicates that the ZINB regression model fitted better than ZIP model ($P < 0.0001$). The father's educational level was significant in both parts of the ZINB regression model ($P < 0.05$), implying that the degree of caries experience increases in children whose fathers have a low level of education, while the excess of caries-free children decreases. Moreover, the increase of coefficients in the zero-inflated part of ZINB regression model implies that the excess of caries-free subjects increases with the later age of tooth eruption. The observed underestimation of the frequencies of zero dmfs counts by the Poisson model is a common result when a dual-group process is not taken into account. *Conclusions:* These regression models provide a useful approach to handling count outcomes as dmfs/DMFS index in caries epidemiology.

Key words: caries epidemiology; negative binomial; Poisson; statistical models; zero-inflated models

Solinas Giuliana, Institute of Hygiene and Preventive Medicine, University of Sassari
Via Padre Manzella, 4 I-07100 Sassari, Italy
Tel.: +39 079 229959
Fax: +39 079 228054
e-mail: giuliana.solinas@uniss.it

Submitted 11 September 2008;
accepted 8 August 2009

doi: 10.1111/j.1600-0528.2009.00500.x

The dmfs/DMFS index is commonly used in epidemiological studies on dental caries prevalence and experience; it is a simple addition of the number of decayed, missing or filled teeth (or surfaces) and represents the cumulative severity of dental caries experience in an individual or a population. It received remarkably little challenge over the first 50 years of its use (1).

However, dmfs/DMFS has weak points and, recently, proposals for a replacement of this index have been made (2). The World Health Organization (WHO) proposed a new index called the 'Significant Caries Index' (SiC) in order to bring attention to those individuals with the highest caries scores in each population (3). From the 1970s, the prevalence of dental caries showed a decrease in industrialized countries (4–8) and reached a *plateau* after 1980 (9).

Consequently, these changes have had the effect of increasing the proportion of zeroes in the distribution of dmfs/DMFS. Zeroes are outcome values and it is important to explicitly account for them in analysis. Since counting outcomes do not meet the normality assumption that is required by many standard statistical tests, analysts have relied on a transformation to induce normality, which often does not work, or on categorization of the outcome which may result in loss of information. A Poisson distribution might be considered as a useful alternative to counting processes under the condition that the expected value is equal to the variance. However, many counting outcomes exhibit more variability than that described by the nominal variance under the Poisson model, a condition called overdispersion. The consequences can be severe if overdispersion is not considered. There are also other ways to deal with overdispersion, e.g. using a sandwich estimator for the variance. This method takes into account the clustering of the counts (10).

However, analysing dmfs/DMFS datasets is necessary to deal both with zero-inflation and overdispersion.

Modifications to the Poisson distribution have also been proposed in other applications, for example, estimating confidence intervals for incidence rates where the case count may be inflated due to false positives (11).

Statistical models for counting data with an excess of observed zeroes have received some attention in recent years, especially in the econometric literature (12–14). In the context of dental epidemiology, one solution is to assume that there are two latent or unobserved groups which could contribute to the excess zeroes: a subpopulation of children who comes from the zero state (they have no decayed teeth due to their personal characteristics) and another subpopulation of children, who are susceptible to caries development, or zero for chance or misclassification (15). Models that recognize the existence of these two groups, and also allow for covariate adjustment in each group, have been developed and are commonly referred to as zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) models (16–18). The ZINB model takes care of both overdispersion and zero-inflated issues. The Zero part takes care of the zero-inflation and the negative binomial (NB) part takes care of the correlated aspect of the dmfs.

Nevertheless, in literature, some authors still continue to encourage and initiate the application of the GLM Poisson model in the analysis of covariates of dental caries (19).

The aim of this paper was to determine the most appropriate regression model to use in children when assessing caries risk factors in overdispersion conditions and in predicting the probability of 'caries-free' subjects. The NB regression model and both the ZIP and ZINB models were compared with the Poisson regression model.

Methods

The sample

The data presented in this paper were gathered as part of the First National Pathfinder Survey amongst children's oral health in Italy. The data of 5538 children (aged 47.2 ± 3.5 months; 51.4% females) were considered for this study. The study sample represented 1.04% of the total population aged 4 years old attending preschool in Italy. The study design included a dental examination and an *ad hoc* questionnaire; 1194 (21.56%) children were affected by caries disease.

Clinical data on caries disease (dmfs) were collected following standardized methods (20, 21). The questionnaire requested information on the child's social, behavioural, ethnic and demographic status through a series of 33 closed questions which focused on sociodemographic background, oral hygiene behaviour and other information related to caries risk. However, only a few variables from the questionnaire were significantly associated to dmfs. The variables selected from the questionnaire were: ethnicity (Italian versus not Italian), a high level of parental education (upper-school or higher) versus a low level (lower than upper-school), preterm birth (<36 weeks), breast feeding, age of tooth eruption (6 months reference versus higher 6-monthly intervals). These were identified as predictors of caries index in the regression models.

Statistical analysis and procedures

In this study, the distribution of dmfs was nonnormal, highly skewed (Skewness = 7.1) and contained excessive zeroes, compared with standard count distributions. The mean of dmfs index was 1.4 (SD = 5).

Although Poisson distribution is typically used for counting outcomes, it cannot be used when an overdispersion, which occurs when the variance of the distribution is greater than the mean value, is present. This distribution can be closely described by the NB model (22). The overdispersion was tested by the statistical test proposed by Bohning (23):

$$O = \frac{\sqrt{\frac{n-1}{2}}\left(s^2 - \overline{x}\right)}{\overline{x}}$$

Moreover, the appropriateness of the NB regression model was tested by a likelihood-ratio test which examines the value of the parameter alpha. If alpha = 0, the process is Poisson.

The Poisson regression model is defined as

$$P(y_i|x_i) = \frac{e^{-\mu}\mu_i^{y_i}}{y_i!}$$

where $\mu_i = e^{x_i\beta}$, with x_i a vector of predictor variables and β a vector of unknown coefficients to be estimated.

To account for an excess of zeroes, the ZIP model is defined as a mixture of two distributions to incorporate extra zeroes:

$$P(y_i|x_i) = \begin{cases} \pi_i + (1-\pi_i)e^{-\mu_i} & y_i = 0 \\ (1-\pi_i)\frac{e^{-\mu_i}\mu_i^{y_i}}{y_i} & y_i > 0 \end{cases} \tag{1}$$

where π_i is the probability of being an extra zero and μ_i can be modelled as $e^{x_i\beta}$ and π_i as $g(z_i\xi)$ where x_i and z_i are vector of predictors which can be different and $g(.)$ is a logit function. The ZINB is formulated as [1], replacing the Poisson distribution $e^{-\mu}\mu_i^{y_i}/y_i!$ with the negative binomial distribution in Equation [2],

$$P(y_i|x_i) = \frac{\Gamma(y_i + 1/\Phi)}{y_i!\Gamma(1/\Phi)} \times \frac{(\Phi\mu_i)^{y_i}}{(1+\Phi\mu_i)^{y_i+1/\Phi}} \tag{2}$$

where Φ is the overdispersion parameter, the ZINB equation [3] is obtained

$$P(y_i|x_i) = \begin{cases} \pi_i + (1-\pi_i)\frac{1}{1+\Phi\mu_i}^{1/\Phi} & y_i = 0 \\ (1-\pi_i)\frac{\Gamma(y_i+1/\Phi)}{y_i!\Gamma(1/\Phi)} \times \frac{(\Phi\mu_i)^{y_i}}{(1+\Phi\mu_i)^{y_i+1/\Phi}} & y_i > 0 \end{cases} \tag{3}$$

which takes into account that the nonzero counts might be correlated (24).

All these regression models were fitted to our data using the POISSON, NBREG, ZIP and ZINB modules of the STATA 9 statistical software (Stata Corp., College Station, TX, USA). The likelihood-ratio test (LR) was considered to compare ZIP to ZINB model, since the ZINB model nests within the ZIP model, when alpha = 0. Based on zero-inflated model, a Vuong nonnested test (V) is then considered to discern the zero-inflated count data model (e.g. either ZIP or ZINB) against its parent distribution (Poisson or NB, respectively), because neither ZIP nor ZINB are nested within its parent distribution, Poisson or Negative Binomial (NB), respectively. The Vuong statistic follows a standard normal distribution; so it can be compared with z-values: when V is greater than 1.96 favours the ZI count model, while V inferior than −1.96 distinctly favours the standard count model; otherwise, neither model is preferred (25). In addition, it was compared the goodness-of-fit of models calibrated using the Akaike information criterion (AIC) (26, 27). The model with the lowest AIC was preferred.

The difference between the observed and predicted probabilities from each model was plotted to obtain graphical illustrations of fit. The predicted probabilities were constructed according to Long and Freese's (28) approach and were adjusted for all covariates. Outcome variable in the regression models was the severity of the disease (dmfs index), while the statistically significant variables on bivariate analysis (Table 1) were considered. The independence between categorical variables was tested with χ^2-test. The inclusion criterion for variables was set up at 10%.

Results

The results of associations between dmfs (categorized as caries experience) and predictor variables

Table 1. Association between dmfs and predictor variables

Variable	dmfs = 0 *n* (%)	dmfs > 0 *n* (%)	*P*-value[a]
Mother's nationality			
Italian	3827 (80.7)	917 (19.3)	<0.001
Not Italian	376 (63.8)	213 (36.2)	
Father's nationality			
Italian	3877 (80.9)	918 (19.1)	<0.001
Not Italian	281 (58.9)	196 (41.1)	
Mother's education			
High	2945 (81.8)	657 (18.2)	<0.001
Low	1240 (72.1)	480 (27.9)	
Father's education			
High	2642 (82.4)	565 (17.6)	<0.001
Low	1516 (73.2)	554 (26.8)	
Preterm birth			
Yes	397 (74.5)	136 (25.5)	0.011
No	3757 (79.2)	985 (20.8)	
Breast feeding			
Yes	3415 (79.3)	891 (20.7)	0.038
No	768 (76.3)	238 (23.7)	
Age of tooth eruption			
<6 months	1226 (75.6)	395 (24.4)	0.004
6–9 months	2013 (79.5)	518 (20.5)	
9–12 months	697 (80.7)	167 (19.3)	
>12 months	163 (81.9)	36 (18.1)	

The total number of children in each characteristic may differ because of missing data.
[a]χ^2-test.

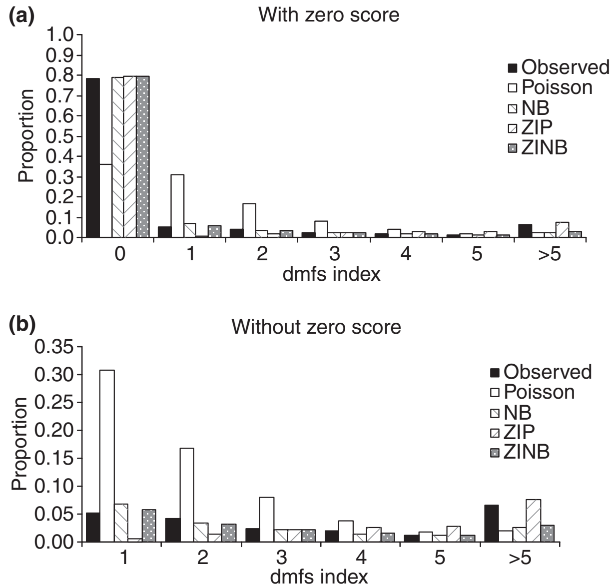

Fig. 1. Distribution of dmfs index according to observed and predicted proportions from four regression models. (a) with zero score, (b) without zero score.

are shown in Table 1. Two groups of children with dmfs = 0 and dmfs > 0 were considered according to the significant variables, as resulted by the chi square test ($P < 0.05$). Significant association between caries experience and potential risk factors were only observed for parents' nationality ($P < 0.001$), parental educational level ($P < 0.001$), preterm birth ($P = 0.011$), prolonged breast feeding ($P = 0.038$) and early tooth eruption ($P = 0.004$).

Model results and comparisons

The Poisson model showed a poor fit. As regard goodness-of-fit testing, the overdispersion statistical test was statistically significant (O = 141.06, $P < 0.0001$). When comparing the NB to the Poisson, the likelihood-ratio test gave a highly significant chi square value ($\chi^2 = 18397.92$, $P < 0.0001$) implying that NB is favoured for this dataset. Moreover, the Vuong statistics of 16.25 indicated that the ZIP model fits better than the Poisson model. Consequently, the expected probability based on the ZIP and NB models showed considerable improvement in the model fitting. Moreover, the ZINB model fits better than NB model ($V = 4.01$, $P < 0.0001$). Furthermore, considering the AIC values for the regression models (Poisson: 5.84; ZIP: 3.18; NB: 2.09 and ZINB: 2.08), it could be observed that the ZINB model fits the data slightly better than the other models.

Figure 1a shows the observed proportions of dmfs index from the four regression models compared with those observed in the sample. The major failure of the Poisson regression model is in predicting the number of zeroes with an underprediction of about 0.5, while at count dmfs = 1, there is an over prediction of about 0.3. The NB and related ZI model are virtually indistinguishable on the plot 1a due to of the excess of zeroes which flats the distributions for dmfs count ≥1, and both fit the number of zeroes quite well. The ZIP model underpredicts individuals with dmfs from one to two (Fig. 1b); conversely, the NB model slightly overpredicts count of dmfs = 1. Overall, the ZINB model makes better predictions than the other models. The high likelihood ratio statistics (5375.43) used to compare ZIP and ZINB model indicates that the ZINB model provides a significant improvement in the fit of the ZIP model ($P < 0.0001$).

The results of fitting the ZINB regression model to the dmfs index with the disease's predictors are reported in Table 2. The adjusted dmfs, obtained from the coefficients of the negative binomial process of the ZINB model, was statistically significant for father's nationality and educational

Table 2. Comparison between GLM Poisson and ZINB models

	GLM Poisson			ZINB: Negative binomial part				ZINB: Zero-inflated part			ZINB: Probability of being an extra zero	
	b	SE	*P*-value	*b*	SE	*P*-value	Adjusted dmfs	*b*	SE	*P*-value	*P*	95% CI
Intercept	2.27	0.04	<0.001	2.29	0.22	<0.001	9.9	−2.05	0.52	<0.001	0.13	0.05–0.36
Nationality												
Italian mother	−0.63	0.03	<0.001	−0.14	0.35	0.70	8.6	0.55	0.45	0.225	0.22	0.09–0.54
Italian father	−0.56	0.03	<0.001	−0.81	0.36	0.02	4.4	0.95	0.49	0.054	0.33	0.13–0.86
Education												
High, mother	−0.33	0.04	<0.001	−0.61	0.14	<0.001	5.4	0.13	0.16	0.39	0.15	0.11–0.20
High, father	−1.07	0.04	<0.001	−0.33	0.14	0.02	7.1	0.52	0.16	0.001	0.22	0.16–0.30
Preterm birth	0.18	0.04	<0.001	0.07	0.18	0.69	10.6	−0.11	0.22	0.613	0.12	0.07–0.18
Breast feeding	−0.11	0.03	<0.001	−0.01	0.13	0.92	9.8	0.22	0.16	0.19	0.16	0.12–0.22
Age of tooth eruption (months)												
6–9 versus 6	−0.15	0.03	<0.001	−0.02	0.12	0.85	9.6	0.26	0.15	0.09	0.17	0.12–0.22
9–12 versus 6	−0.15	0.04	<0.001	0.05	0.17	0.78	10.4	0.57	0.19	0.004	0.23	0.15–0.33
>12 versus 6	0.04	0.06	0.529	0.36	0.31	0.24	14.2	0.77	0.31	0.015	0.28	0.15–0.51

level and for mother's educational level. The lower values, compared to the intercept one, indicate the 'protective' effect of these variables. In particular, the educational level of the father was significant in both parts of the ZINB regression model, implying that the degree of caries experience increases for the group of children whose father's educational level is low, while the excess of the caries-free children decreases. The probability of being an extra zero, or 'caries-free', reported in the twelfth column in Table 2, could be interpreted as the estimate for an individual who is in the group of children whose fathers have a high educational level, and whose *P*-value is 0.22. Moreover, the coefficients for six-monthly age intervals of tooth eruption higher than 6 months versus 6 months imply that the higher the age of first tooth eruption, the lower the probability of being affected by caries. Therefore, the probability of being an extra zero increases with the age of tooth eruption, ranging from 0.17 in children with tooth eruption at 6–9 months to 0.28 in children with tooth eruption at over 12 months.

Discussion and conclusion

The process of choosing the best model to examine risk factors which are associated with a disease is a trade-off between simplicity and accuracy. This is particularly true for caries disease in childhood since a large proportion of children are caries-free (zero counts) according to dmfs index, while a small number of children typically account for an extreme amount of caries. Although the Poisson regression model is still recommended for analysing count data, it almost always does not fit very well because of overdispersion. Moreover, it is important to mention that the Poisson model must be used for modelling independent counts but dmfs/DMFS is a count based on dependent counts. In fact, most count data are detected in the same mouth. Consequently, when the estimated model is used to predict the probability of an event, i.e. in this study the dependence of dmfs index on the influence of childhood sociodemographic factors and on the common risk factors linked to some prenatal and postnatal conditions (preterm of birth, age of tooth eruption, and breast feeding) (29–31), there is the risk of biased prediction. In particular, in this study, the Poisson regression model resulted a poorly fitted model, underpredicted the observed number with zero dmfs values and overpredicted the number of subjects with one, two or three events. This, in turn, may lead to spurious conclusions, such as concluding that a factor is important in predicting caries disease when in fact this may not be the case.

The standard Poisson model is not suitable when the variance exceeds the mean, nor when the events are dependent. Extensions of the Poisson model – incorporating an overdispersion parameter, or using generalized estimating equations (GEE) methods when a fairly large number of small clusters is present or a longitudinal study is considered, or, finally, the use of the negative binomial distribution and/or zero-inflated models – are now available in many software packages and address many of the shortcomings of the overly-simplistic Poisson model. Each of the above extensions could be applied according to the available kind of dataset. Therefore, to correctly estimate the covariates of dental caries beside the traditional regression models (32, 33) a large variety of new models like the zero-inflated were brought to the attention of a wider epidemiological audience.

In the present study, which was a cross-sectional survey, where greater differences in the dmfs values have been found amongst large sections than within them, and where the variance highly exceeded the mean, with a very large number of zero count (34) the detected frequency of zero was larger than that predicted by the Poisson regression model. Consequently, a zero-inflated model should be more attractive.

In fact, all factors which have been considered into the Poisson regression model resulted as significant predictors for dmfs index (fourth column of Table 2), while NB, ZIP and ZINB models, which predict better when compared with the Poisson model, gave significant results only for some factors. The results of the ZINB regression model showed that the father's educational level was a significant covariate affecting the probability of being caries-free, as well as the delay of tooth eruption after 6 months which increased the probability of being caries-free.

The ZINB model best fitted these data and might provide an appealing tool for explaining the preponderance of zeroes. More than three quarters of the examined sample showed a dmfs = 0. These data are consistent with previous data recorded in European countries (35, 36).

In settings where there is excess of zeroes, zero-inflation models are the best choice and the results of the present study suggest that the Poisson model will no longer be considered for use in cross-sectional study and ZINB will be preferred if high variance and number of an excess of zeroes are present. Moreover, as reported elsewhere, the ZINB approach was found to have the best fit not only with cross-sectional deciduous caries data, but also with permanent dentition data in longitudinal study (16). One advantage of these models is that they can estimate the probability of being a zero as a function of covariates, rather than the degree of caries experience.

In this setting, the implementation of *ad hoc* methods might be useful in order to estimate both the amount of dmfs = 0 amongst all the subjects, including zero and non zero values, and, in a separate model, the proportion of subjects in a zero state (i.e. caries-free). Each zero-inflated model retains the underlying assumption that a dual-group process may be operated. So, even if they fit data better than the conventional models, complexity and difficulty are introduced in interpreting the modelling results. For the dual-group process, one group will not experience caries (zero state), while the second group is a caries state, i.e. those that most likely show caries, but by chance are caries-free at the moment of observation, for whom the frequencies of counts follow some known distribution (either Poisson or negative binomial). To deal with dual-group phenomenon, statistical procedures able to recognize the two groups were used in order to improve goodness-of-fit; the use of ZIP and ZINB, which assume that some initial process brings children into an at-risk population, allows for this, as well as for the possibility of detecting that these two groups may be influenced by different factors in different ways.

Using the Vuong's test to determine if the ZIP model achieved a better fit to the data than the Poisson model, we argued that the presence of a dual-group phenomenon was in operation. Furthermore, the ZINB model was the best one to use for this dataset, even though the goodness-of-fit plot indicated no visual difference between NB and ZINB prediction, because of the excess of zeroes that made the distribution of dmfs ≥1 counts flat. The underprediction of the frequencies of zero dmfs counts by the Poisson model, as well as overprediction of the frequencies of children who had dmfs from 1 to 3, are typical problems encountered when a dual-group process is not taken into account. The zero-inflated models have been consistently judged to be very useful because they efficiently fit the data and can model the overall structure of the data. Moreover, they provide a means for covariate adjustment and risk factors assessment and are not difficult to fit using common statistical software.

Statisticians and epidemiologists must analyse their data and utilize models which provide an appropriate fit and meaningful interpretation. In particular, for models of dental caries epidemiology, attention should be paid to the functional form of the outcome to ensure that underlying assumptions of the utilized methods are met.

References

1. Burt BA, Eklund SA Dentistry, dental practice and the community, 4th edn. Philadelphia: WB Saunders Co, 1992.
2. Spencer AJ. Skewed distributions – new outcome measures. Community Dent Oral Epidemiol 1997;25:52–9.
3. Bratthall D. Introducing the Significant Caries Index together with a proposal for a new global oral health goal for 12-year-olds. Int Dent J 2000;50:378–84.
4. Frencken JE, Kalsbeek H, Verrips GH. Has the decline in dental caries been halted? Changes in caries prevalence amongst 6- and 12-year-old children in Friesland, 1973–1988 Int Dent J 1990;40: 225–30.
5. Truin GJ, König KG, de Vries HCB, Mulder J, Plasschaert AJM. Trends in caries prevalence in 5-, 7- and 11-year-old schoolchildren in The Hague between 1969 and 1989. Caries Res 1991;25:462–67.
6. Nugent ZJ, Pitts NB. Patterns of change and results overview 1985/6–1995/6 from the British Association for the Study of Community Dentistry (BASCD) coordinated National Health Service surveys of caries prevalence. Community Dent Health 1997;14:30–54.
7. Downer MC. The changing pattern of dental disease over 50 years. Br Dent J 1998;185:36–41.
8. Truin GJ, König KG, Bronkhorst EM, Frankenmolen F, Mulder J, van't Hof MA. Time trends in caries experience of 6- and 12-year-old children of different socioeconomic status in The Hague. Caries Res 1998;32:1–4.
9. Truin GJ, Konig KG, Kalsbeek H. Trends in dental caries in the Netherlands. Adv Dent Res 1993;7:15–18.
10. Petersen MR, Deddens JA. A comparison of two methods for estimating prevalence ratios. BMC Med Res Methodol 2008;8:9.
11. Kegler SR. Reporting incidence from a surveillance system with an operational case definition of unknown predictive value positive. Epidemiol Persp Innovation 2005;2:7.
12. Greene WH Accounting for excess zeroes and sample selection in Poisson and negative binomial regression models. Working paper, New York: Department of Economics, Stern School of business, New York University, 1994.
13. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 1992;34:1–14.
14. Mullahy J. Heterogeneity, excess zeros, and the structure of count data models. J Appl Economet 1997;12:337–50.
15. Böhning D, Dietz E, Schlattmann P, Mendonça L, Kirchner U. The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. J Royal Stat Soc A 1999;162:195–209.
16. Lewsey JD, Thomson WM. The utility of the zero-inflated Poisson and zero-inflated negative binomial models: a case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. Community Dent Oral Epidemiol 2004;32:183–89.
17. Mwalili SM, Lesaffre E, Declerck D. The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research. Stat Methods Med Res 2008;17:123–39.
18. Hur K, Hedeker D, Henderson W, Khuri S, Daley J. Modeling clustered count data with excess zeros in health care outcomes research. Health Serv Outcomes Res Methodol 2002;3:5–20.
19. Javali SB, Parameshwar VP. Use of the generalized linear models in data related to dental caries index. Indian J Dental Res, 2007;18:163–7.
20. Campus G, Solinas G, Cagetti MG, Senna A, Minelli L, Majori S et al. National pathfinder survey of 12-year-old children's oral health in Italy. Caries Res 2007;41:512–17.
21. Castiglia P, Campus G, Solinas G, Maida C, Strohmenger L. Children's oral health in Italy. Training and clinical calibration of the examiners for the National pathfinder about caries disease. Oral Health Prev Dent 2007;5:255–61.
22. Grainger RM, Reid DB. Distribution of dental caries in children. J Dent Res 1954;33:613–23.
23. Böhning D, Dietz E, Schlattmann P Zero-inflated Count models and their applications in public health and social science. In: Rost J, Langeheine R eds. Applications of latent trait and latent class models in the Social Sciences. 1st edn. Germany: Munster Waxmann,1997; 333–44.
24. Cameron AC, Trivedi PK. Regression analysis of count data. New York: Cambridge University Press; 1998.
25. Vuong Q. Likelihood ratio tests for model selection and non-nested hypothesis. Econometrica 1989;57:307–33.
26. Akaike H Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F eds., 2nd international symposium on information theory. Budapest: Akademiai Kiado, 1973.
27. Akaike H. A new look at the statistical model identification. IEEE Automat Contr, 1974;6:716–23.
28. Long JS, Freese J. Predicted probabilities for count models. The Stata J 2001;1:51–7.
29. Diane L, Eastman MA. Dental outcomes of preterm infants. Newborn Infant Nurs Rev 2003;3:93–98.
30. Alvarez JO, Navia JM. Nutritional status, tooth eruption, and dental caries: a review. Am J Clin Nutr 1989;49:417–26.
31. Iida H, Auinger P, Billings RJ, Weitzman M. Association between infant breastfeeding and early childhood caries in the United States. Pediatrics 2007;120:e944–e952.
32. Fabien V, Obry-Musset AM, Hedelin G, Cahen PM. Use of the generalised linear model with Poisson distribution to compare caries indices. Community Dent Health 1999;16:93–6.
33. Stecksén-Blicks C, Sunnegårdh K, Borssén E. Caries experience and background factors in 4-year-old

children: time trends 1967–2002. Caries Res 2004;38:149–55.
34. Campus G, Solinas G, Strohmenger L, Cagetti MG, Senna A, Minelli L et al. National Pathfinder Survey on children's oral health in Italy. Pattern and severity of caries disease in 4-year-olds. Caries Res, 2009;43:155–62.
35. Henkuzena I, Care R, Rogovska I. Dental Status Among 2–6 year old children in Riga City, Latvia. Stomatologija. Baltic Dental Maxillofac J 2004;6: 28–30.
36. Olak J, Mändar R, Karjalainen S, Söderling E, Saag M. Dental health and oral mutans streptococci in 2–4-year-old Estonian children. Int J Paediatr Dent 2007;17:92–7.