Caries Res 2012;46:424–426

DOI: [10.1159/000339374](https://doi.org/10.1159/000339374)**Statistical Analysis of Dental Caries: Different Methods for Different Outcomes**D.L. Long^a, J.S. Preisser^a, J. Stamm^b^aDepartment of Biostatistics, Gillings School of Global Public Health, University of North Carolina, and ^bDepartment of Dental Ecology, School of Dentistry, University of North Carolina, Chapel Hill, N.C., USA

Sir,

The recent paper by Campus et al. [2011] explored a possible relationship between smoking levels and dental caries experience in 21- to 32-year-old adults attending an Italian military academy. We report some reservations and concerns that several aspects of the study may be deficient on methodological grounds. Of primary importance is how the outcome of interest, decayed permanent tooth surfaces (DS), is categorized into an ordered categorical variable for the authors' chosen analyses. As seen in table 5 of their article, the authors classified each individual into one of three categories of DS (0, 1–3, >3), 'using as reference the categorization of DMFS by Lesaffre et al. [2004]'. But in examining the methodology of Lesaffre et al., we find that they did not consider DS or DMFS scores, nor did their study even deal with permanent teeth. Rather, Lesaffre et al. considered only grade 1 children and ordered their outcome variable (dmft) into one of four contiguous categories.

Second, Campus et al. indicate that they performed a zero-inflated Poisson (ZIP) regression analysis on the categorized DS data. However, a ZIP model for a coarsely categorized outcome is not well justified because data set out into ordinal categories would most likely be analyzed using ordered categorical logistic regression as described by Mwalili et al. [2005] and Preisser et al. [2011], and implemented by Lesaffre et al. [2004]. As originally described in Lambert [1992], ZIP regression is specifically designed to handle count data with many zeros. While ZIP regression is more commonly being applied to caries count indices (e.g. DMFS, DS, dfmt, etc.), ordered categorizations such as Campus et al. employ require a completely different set of analysis tools.

A third concern regards the distinction made between a light and heavy smoker. The authors define a 'light smoker, when a

subject reports to smoke <10 cigarettes *and* has started the habit <3 years, and heavy smoker when he/she smokes >10 cigarettes a day *and* he/she is smoking for >3 years'. The italics have been added for emphasis to highlight that one of the 'and' should have been written as 'or' so that any smoker in the sample may be categorized as a light or heavy smoker. Clearly, the Campus et al. smoker classification scheme is non-exhaustive, and does not comport with established methods. How, for example, did Campus et al. classify smokers using >10 cigarettes a day, smoking for <3 years, or smokers using <10 cigarettes a day for >3 years?

In addition to these general concerns, there are a number of apparent inconsistencies in the reporting of numerical results. In table 4, the sample size (n = 426) does not match that given in the Results section (n = 436) and the reported odds ratio of 1.88 does not match that calculated from the table entries (1.78); moreover, the row percentages are switched. The reported odds ratios in table 5 are inexplicably grossly inflated with respect to those that can be calculated from the table, and the sample size (n = 765) therein is greater than the reported sample for the study (n = 763). Finally, in table 6, two of the confidence intervals do not contain (as they should) the parameter coefficient estimates.

References

- Campus G, Cagetti MG, Senna A, Blasi G, Mascolo A, Demarchi P, Strohmenger L: Does smoking increase risk for caries? A cross-sectional study in an Italian military academy. *Caries Res* 2011;45:40–46.
- Lambert D: Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 1992;34:1–14.
- Lesaffre E, Mwalili SM, Declerck D: Analysis of caries experience taking inter-observer bias and variability into account. *J Dent Res* 2004;83:951–955.
- Mwalili SM, Lesaffre E, Declerck D: A bayesian ordinal logistic regression model to correct for interobserver measurement error in a geographical oral health study. *Appl Statist* 2005;54:77–93.
- Preisser JS, Phillips C, Perin J, Schwartz TA: Regression models for patient-reported measures having ordered categories recorded on multiple occasions. *Community Dent Oral Epidemiol* 2011;39:154–163.

Reply

Different Statistical Approach, Same Results

G. Campus^{a,b}, M.G. Cagetti^b, A. Senna^b, G. Blasi^c,
A. Mascolo^c, P. Demarchi^c, L. Strohmer^b

^aDental Institute, University of Sassari, Sassari,

^bWHO Collaborating Centre of Milan for Epidemiology and
Community Dentistry, University of Milan, Milan, and

^cDental Institute, University of Genoa, Genoa, Italy

The relationship between smoking habits and oral health is widely described in the literature. Apart from the well-documented relationship with periodontal diseases, the literature analyzing the correlation between caries and smoking habits is scant [Johnson and Bain, 2000]. In the paper entitled 'Does smoking increase risk for caries? A cross-sectional study in an Italian military academy' [Campus et al., 2011], the authors tried to verify a possible association.

The study was designed as a cross-sectional study and no information about the history of the relation between smoking and caries is given. Data were collected through a self-administered standardized questionnaire and a dental examination. The authors believe that the methodological approach used in the paper is correct; nevertheless, some shortcomings can be raised, as established by D. Leann Long, John Preisser and John Stamm.

The low caries experience and prevalence in people from Western countries had the effect of increasing the proportion of zeroes in the distribution of DMFS and components. Zeroes are outcome values and it is important to account for them explicitly in the analysis. Since counting outcomes does not meet the normality assumption, analysts have relied on a transformation to induce normality, which often does not work, or on categorization of the outcome. 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 over-dispersion. The consequences can be severe if over-dispersion is not considered. In a previous paper, the same research group [Solinas et al., 2009] evaluated the probability of 'caries-free' subjects and the dependence of DMFS index on the influence of childhood socio-demographic factors, through the application of regression models (Poisson, negative binomial and zero-inflated models, Poisson and negative binomial). The zero-inflated Poisson model achieved a better fit to the data than the Poisson model; the zero-inflated negative binomial model was the best one to use for the dataset object of the paper.

In the paper of Campus et al. [2011] the distribution of DS was non-normal, highly skewed and contained excessive zeroes, compared with standard count distributions. The mean of DS index was 0.9 (SE = 0.1) and a high over-dispersion was present (dispersion parameter was significantly different from zero $\alpha = 2.6$, 95% CI = 2.0–3.1), as it occurs when the variance of the distribution is greater than the mean value. To account for an excess of zeroes, the zero-inflated Poisson regression model was used as it is defined as a mixture of two distributions to incorporate extra zeroes. The methodology described by Lesaffre et al. [2004] was used, even if in that paper the deciduous dentition was considered.

Table 1. Ordinal logistic regression for DS categorization

	Coefficient (SE)	p value	95% CI
Gingival status (healthy)	0.58 (0.29)	0.04	0.15; 1.16
Self-satisfaction teeth and gums (yes)	−0.84 (0.17)	<0.01	−1.18; −0.49
Dental check-up (1/6 months)	−0.98 (0.42)	0.01	−1.37; −0.33
Smoking habits (≥3 years)	−0.60 (0.14)	0.02	−0.93; 0.27
Cut 1	0.05 (0.52)		−0.98; 1.08
Cut 2	1.50 (0.53)		0.46; 2.55

Number of observations = 739; log likelihood = −529.05, $p < 0.01$.
Missing data due to no responders to questionnaire items = 23.

Table 2. Output from the log of Stata used for table 5

Bleeding	Cases	Controls	Odds	95% CI
ds 0	0	487	79	
ds 1–3	1	125	13	9.61538 5.43142–17.02237
ds >3	2	58	3	19.33333 6.05781–61.70177

Score test for trend of odds: $\chi^2(1) = 5.52$. $\text{Pr} > \chi^2 = 0.0188$.

Table 3. Zero-inflated Poisson regression-adjusted estimates for DS categorization

	Coefficient (SE)	p value	95% CI
Gingival status (healthy)	0.51 (0.25)	0.04	0.02; 1.01
Self-satisfaction teeth and gums (yes)	−0.65 (0.14)	<0.01	−0.92; −0.38
Dental check-up (1/6 months)	−0.70 (0.28)	0.01	−1.25; −0.14
Smoking habits (≥3 years)	−0.34 (0.14)	0.02	−0.65; 0.24
Constant	−0.11 (0.39)	0.77	−0.88; 0.65

Number of observations = 739, log likelihood = −541.74, $p < 0.01$.
Missing data due to no responders to questionnaire items = 23.

Several methodological and statistical approaches might be used to analyze caries data. The approach proposed by D. Leann Long, John Preisser and John Stamm is appealing. Following the suggestion, an ordinal logistic regression on the dataset was performed. It is noteworthy that the same covariates were statistically associated to the categories of DS: gingival status (healthy), self-satisfaction teeth and gums (yes), dental check-up (1/6 months), smoking habits (≥3 years) (table 1).

During the review process, in order to simplify data analysis, it was decided to categorize the sample positive to smoking in two groups only: light and heavy smokers. Unfortunately, in the last draft of the manuscript in table 1 the replacement of 'and' with 'or' was missed.

The number of heavy smokers (>3 years) was 436 as reported in the Results section but 10 of them did not declare for how many years they were exactly smoking and so they were excluded from table 4, as reported at the bottom of the table. Regrettably, a switch of row percentage occurred.

Regarding table 5, data were obtained from the output of Stata, that is now displayed in table 2. In table 6 the confidence intervals for self-satisfaction teeth and gums and smoking habits were erroneously reported; the correct terms are displayed in table 3.

In conclusion, the authors can affirm that even though some small slipups occurred, the results did not suffer and the study was performed using a correct methodology. The comparison with an ordinal logistic regression procedure, as suggested by D. Leann Long, John Preisser and John Stamm, confirms it, showing comparable results and assessing the validity of the two different methodological approaches.

Since the role of tobacco smoking is still uncertain when it comes to the development of dental caries, it cannot be assumed to be a major risk factor for this disease and therefore justifies related community and clinical preventive strategies.

References

- Campus G, Cagetti MG, Senna A, Blasi G, Mascolo A, Demarchi P, Strohmenger L: Does smoking increase risk for caries? A cross-sectional study in an Italian military academy. *Caries Res* 2011;45:40–46.
- Johnson NW, Bain CA: Tobacco and oral disease. EU Working Group on Tobacco and Oral Health. *Br Dent J* 2000;189:200–206.
- Mwalili SM, Lesaffre E, Declerck D: A bayesian ordinal logistic regression model to correct for interobserver measurement error in a geographical oral health study. *Appl Stat* 2005;51:77–93.
- Preisser JS, Phillips C, Perin J, Schwartz TA: Regression models for patient-reported measures having ordered categories recorded on multiple occasions. *Community Dent Oral Epidemiol* 2011;39:154–163.
- 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.

Dr. Guglielmo Campus
Dental Institute, University of Sassari
Viale San Pietro 43/C
IT-07100 Sassari (Italy)
Tel. +39 079 228 540, Fax +39 079 228 541, E-Mail gcampus@uniss.it