Logistic regression analysis of oral health data in assessing the therapeutic value of amine fluoride containing products

Sándor Márton[1], Gábor Nagy[2], Katalin Gábris[3], Jolán Bánóczy[4], Melinda Madléna[3]

[1] Department of Sociology, University of Debrecen, Debrecen. [2] Department of Oral Diagnostics, Semmelweis University, Budapest, Hungary. [3] Department of Paediatric Dentistry and Orthodontics, Semmelweis University, Budapest, Hungary. [4] Department of Oral Biology, Semmelweis University, Budapest, Hungary

Abstract

Aims. One aim of the study was to evaluate the impact of using logistic regression in oral health research and to evaluate the effects of amine fluoride (AmF) products in remineralization, statistically. ***Methods.*** The longitudinal clinical study evaluated the effect of combined use of AmF toothpaste and gel (test group 1) versus using AmF toothpaste and placebo gel (test group 2), and material containing no AmF at all (control group). Reversal of white spot lesions was examined using logistic regression. ***Results.*** The use of AmF toothpaste combined with AmF gel reduced the number of incipient lesions significantly ($p<0.001$) compared to the group using no AmF products. Test group 1 had 2.3 times higher chance of remineralization of incipient lesions than test group 2 ($p = 0.03$), and 11.1 times higher chance (p 0.001), than the control group. The significant odds ratio showed that the combined use of AmF gel and toothpaste was more effective in remineralizing incipient lesions. ***Conclusions.*** Logistic regression can be used to predict a dependent variable on the basis of independent variable and to determine the proportion of variance in the dependent variable explained by the independent variables, to rank the relative importance of independent variable, to assess interaction effects and to understand the impact of covariate control variables. This method could detect the beneficial effects of AmF products not only clinically but also statistically.

Key Words: Amine fluoride/stannous fluoride, logistic regression, reversal of white spot lesion, parameter estimates, odds ratio.
Short title: Logistic regression analysis of data

Introduction

In a previous clinical trial we elaborated a preventive programme, based on the combined use of amine fluoride (AmF) toothpaste and AmF gel. It has been shown that the combined use of AmF products provided a significant reduction of DMFS values both including and not including the incipient lesions [1]. It would be of interest to know if the DMFS reduction is related to the reversal of white spot lesions due to the remineralizing effect of AmF products. In order to assess this problem not only clinically, but also statistically, the logistic regression model can be used to determine this relationship.

Logistic regression is a useful statistical method if the dependent variable has only two possible values, - for example 0 and 1 -, when other methods such as multiple regression become invalid because predicted values of y would not be constrained to range between 0 and 1. Discriminant analysis could also be used in similar circumstances. However, discriminant analysis will only produce optimal solutions if its assumptions are supported by the data. An alternative approach is logistic regression. In logistic regression the dependent variable (**y**) is the probability that an event will occur, hence y is constrained between 0 and 1. Logistic regression has the additional advantage that all of the predictors can be binary, a mixture of categorical and continuous or just continuous [2,3].

The aim of this article was to help to better understand the use of logistic regression in oral

Corresponding author: Dr. Melinda Madléna, Department of Paediatric Dentistry and Orthodontics, Faculty of Dentistry, Semmelweis University, Szentkirályi utca 47. sz.., 1088 Budapest, Hungary. Tel.:361-318-7187, Fax: 361-266-8269, e-mail: madlena@fok.usn.hu

health research and to evaluate the effects of AmF products in remineralization, statistically.

Methods

Design of the study

This study was based on the combined use of amine fluoride (AmF) toothpaste and AmF gel (test group 1), versus the use of AmF toothpaste and placebo gel (test group 2), and material containing no AmF at all (control group). During the clinical trial caries experience, oral hygiene, salivary microbiological factors, buffer capacity were evaluated at baseline, and in the first and then in the second year. The design of the two years longitudinal clinical study is shown in *Table 1.*

Study sample:

Test group 1: AmF gel + toothpaste
Test group 2: AmF toothpaste + placebo gel
Control group: no AmF products

The unit of observation was a *subject's tooth surface* and the dependency problem was the *reversal of incipient lesions* (in otherwise healthy subjects).

During the preparation of the dataset for model fitting, the original caries score of the surface of teeth was recoded during each examination, and received either the code 0 (healthy) or the code 1 (with white spot lesion). At baseline altogether 428 surfaces were selected with white spot lesion of which changes were evaluated mathematically during the study.

The logistic model is written as:

Prob(event) = $\frac{1}{1}$ $Prob(event) = \frac{1}{1+e^{-z}}$ + e^{-1}

where z is $b_0 + b_1x_1 + b_2x_2 + ... + b_Px_P$,

In this model x_1, x_2, x_3, …., x_P are the independent variables. The variable **z** is a measure of the total contribution of all the risk factors used and is known as the logit. The variable **z** is usually defined as $b_0 + b_1x_1 + b_2x_2 + ... + b_Px_P$, where b_0 is called the „intercept" and b_1, b_2, b_3, and so on, are called the „regression coefficients" of x_1, x_2, x_3 respectively. The intercept is the value of z when the value of all risk factors is zero (i.e., the value of **z** in someone with no risk factors).

The logit coefficients, also called nonstandardized *logistic regression coefficients* or *effect coefficients*, correspond to the b (nonstandardised regression) coefficients in ordinary least squares (OLS) regression. Logits are the natural log of the odds. The odds ratio is a measure of effect size and is particularly important in Bayesian statistics and logistic regression. It is defined as the ratio of the odds (possibility) of an event occurring in one group tothe odds of it occurring in another group, or to a sample-based estimate of that ratio. In our study it shows the relative possibility of remineralization of white spot lesions in group using both AmF containing toothpaste and gel compared to the control group using only toothpaste. These are used in the logistic regression equation to estimate (predict) the log odds that the dependent equals 1 (binomial logistic regression) or that the dependent equals its highest/last value (multinomial logistic regression). For the dichotomous case, if the logit for a given independent variable is b_1, then a unit increase in the independent variable is associated with a b_1 change in the log odds of the dependent variable (the natural log of the probability that the dependent = 1 divided by the probability that the dependent = 0). In multinomial logistic analysis, where the dependent may have more than the usual either 0-

Table 1. *Design of the study*

Two years longitudinal clinical study		
n= 586		n=406
Baseline	1 year	2 years
→		
Examined parameters:	**Clinical**:	**DMFS, VPI**
	Microbiological:	**SM, LB counts**
	Biochemical:	**buffer capacity**
	Demographic:	**gender, social status, previous fluoride exposure**

or-1 values, the comparison is always made with the last value rather than with the value of 1. Note that OLS has an identity link function while logistic regression has a logit link function (that is, logistic regression unlike calculates changes in the log odds of the dependent, rather than changes in the dependent itself) [4]. In SPSS output, the logit coefficients are labelled B.

Analysis of data

The sample was taken on the basis of a longitudinal study, where the measurements were made three times and only the values for those subjects, who participated in all three examinations were evaluated**.** Descriptive statistics were calculated for variables to compare the categorical variables. Chi-square statistics and logistic regression were used. All statistical analyses were performed using by SPSS for Windows statistical software version 11.0.

Results

In the control group the constant in the equation was:

$$b_{0=} - 4.06 = \ln (4/233).$$

The value of b1 in Test group 1 was:

$$-1.56-(-4.06)=2.41= b_1$$

and the Odds ratio, (shown as „Exp(B)“ in the „Variables in the Equation“ table) , was

$$Exp(B)=e^{2.41}=11.121.$$

In Test group 2 (AmF toothpaste and placebo gel) these values were:

$$-2.5-(-4.06)=1.56= b_2,$$

$$Exp(B)=e^{1.56}= 4.76, \text{ respectively.}$$

The continued use of AmF products (gel + toothpaste) as judged by parameter estimates was found to be non significant after one year, but it was significant at the end of the second year ($p < 0.001$).

The results showed that the group using AmF gel and toothpaste had 2.3 times higher chance for remineralization of incipient lesions than the group using toothpaste only, ($p = 0.03$), and 11.1 times higher chance ($p < 0.001$), than the group using no AmF products, but maintaining their usual oral care habits. The odds ratio was 4.7 in the group using AmF toothpaste only ($p = 0.009$). The results were statistically significant only in the case of buccal surfaces *(Table 2).*

No significant relationship was found between other clinical, microbiological, biochemical parameters on the one hand and reduction of incipient lesions, on the other.

Discussion

Logistic regression can be used to predict a dependent variable on the basis of independent variable and to determine the percentage of variance in the dependent variable explained by the independent variable; to rank the relative importance of independent variable; to assess interaction effects; and to understand the impact of covariate control variables.

Logistic regression applies maximum likelihood estimation after transforming the dependent

Table 2. *Loglinear model for the relationship between groups and re-mineralizing buccal surfaces*

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	259,540	,055	,131

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95,0% C.I.for EXP(B)	
							Lower	Upper
Group			21,148	2	,000			
Test group 1	2,409	,546	19,492	1	,000	11,121	3,817	32,403
Test group 2	1,560	,601	6,730	1	,009	4,761	1,464	15,475
Constant	-4,062	,504	65,064	1	,000	,017		

into a logit variable (the natural log of the odds of the dependent if present). In this way, logistic regression estimates the probability of the occurrence of a certain event. Logistic regression analogous to OLS regression in several aspects: logit coefficients correspond to b coefficients in the logistic regression equation, the standardized logit coefficients correspond to beta weights, and a pseudo R^2 statistics is available to summarize the strength of the relationship. Also, goodness-of-fit tests such as model chi-square are available as indicators of model appropriateness as is the Wald statistic to test the significance of individual independent variables. Logistic regression applies maximum likelihood estimation after transforming the dependent into a logit variable (the natural log of the odds of the occurrence of the dependent). In this way, logistic regression estimates the probability of occurrence of a certain event.

Log-linear, logit, and probit models extend the principles of generalized linear models (for example regression) to better treat the case of dichotomous and polytomous dependent variables. As such they are a kind of multi-way frequency analysis (MFA), but sometimes labelled as log-linear analysis. These methods differ from standard regression in substituting maximum likelihood estimation of a link function of the dependent for by using regression analysis for the least squares estimation of the dependent. The function used in log-linear analysis is the log of the dependent, y. The function used in logit is the natural log of the odds ratio. The function used in probit is the inverse of the standard normal cumulative distribution function [5,6,7].

In our study the parameter estimates for AmF products were found significant after the second year. The significant odds ratio in the second test showed that the combined use of AmF gel and toothpaste was more effective in remineralizing incipient lesions compared to the other δ groups. Using the logistic regression model showed the advantage for results that it was impressive, but prediction using classification techniques presented high levels of sensitivity [8, 9].

Using the statistical method of this logistic regression it became possible to further improve the effectiveness and scientific value of a clinical study yielding further results based on the same database.

Acknowledgements

The authors gratefully acknowledge the support of GABA International Ltd. (Switzerland) and Hungarian Scientific Grants (T 034389, T 42480).

References

1. Madléna M, Nagy G, Gábris K, Márton S, Keszthelyi G, Bánóczy J. Effect of Amine fluoride Toothpaste and Gel in High Risk Groups of Hungarian Adolescents: Results of a Longitudinal Study. *Caries Res* 2002; **36:** 142-6.
2. Agresti A. Categorial data analysis. New York: John Wiley Publisher, 1990.
3. Walsh A. Statistics for the social sciences with computer applications. New York: Harper & Row Publishers, 1990.
4. Aldrich JH and Nelson FE. Linear probability, logit and probit models. Sage University Papers on Quantitative Applications in the Social Sciences, Sage:CA 1984.
5. Nelder J and Weddenburn RWM. Generalized Linear Models. Statis Soc 1972; **135:** 370-84.
6. Huntington E. Increasing the sensitivity of caries clinical trials by applying logistic modelling to specific teeth. *Community Dent Oral Epidemiol* 1985; **13:** 264-7.
7. Lipsitr SR, Dear KBG, Zhao L. Jackknife estimators of variance for parameter estimates from estimating equations with applications to clustered survival data. *Biometrics* 1994; **50:** 842-6.
8. Beck JD, Lawrence HP, Koch GG. Analytic approaches to longitudinal caries data in adults. *Community Dent Oral Epidemiol* 1997; **25:** 42-51.
9. O'Mullane DM, Kavanagh D, Ellwood RP, Chester RK, Schafer F, Huntington E et al. A three-year clinical trial of a combination of trimetaphosphate and sodium fluoride in silica toothpastes. *J Dent Res* 1997; **76:** 1776-81.

***