Evaluation of clinical dental variables to build classifiers

Silvia Mina

Ana Isabel Azcurra

Carolina Riga

Lila Susana Cornejo

Mabel Brunotto

Medicina Oral, Patología Oral y Cirugía Bucal. Vol. 13, No. 7 (2008), pp. 398-402

http://www.medicinaoral.com/medoralfree01/v13i7/medoralv13i7p398.pdf

Evaluation of clinical dental variables to build classifiers to predict celiac disease

Silvia Mina [1], Ana Isabel Azcurra [2], Carolina Riga [3], Lila Susana Cornejo [2], Mabel Brunotto [4]

(1) Assistant Professor. Social and Preventive Department. Faculty of Dentistry. National Universty of Cordoba-Argentina
(2) Regular Professor. Buccal Biology Department.Faculty of Dentistry. National Universty of Cordoba-Argentina
(3) Specialist Physician. Gastroenterology Office of Children's Hospital "Santísima Trinidad", Córdoba-Argentina
(4) Assistant Professor.Buccal Biology Department.Faculty of Dentistry. National Universty of Cordoba -Argentina

Correspondence:
Dr. MSc. Mabel N Brunotto
Departamento Biología Bucal
Facultad de Odontología
Universidad Nacional de Córdoba
Haya de La Torre s/n Ciudad Universitaria
Córdoba-Capital. Argentina. CP 5000.
E-mail: mbruno@odo.unc.edu.ar

Received: 22/06/2007
Accepted: 09/12/2007

Indexed in:
-Index Medicus / MEDLINE / PubMed
-EMBASE, Excerpta Medica
-SCOPUS
-Indice Médico Español
-IBECS

Mina S, Azcurra AI, Riga C, Cornejo LS, Brunotto M. Evaluation of clinical dental variables to build classifiers to predict celiac disease. Med Oral Patol Oral Cir Bucal. 2008 Jul 1;13(7):E398-402.

http://www.medicinaoral.com/medoralfree01/v13i7/medoralv13i7p398.pdf

Abstract

Objective: The aim of this study was to evaluate the use of salivary variables to build statistical models for predicting celiac disease in symptomatic children. Materials and Methods: the study group consisted of 52 children with celiac disease diagnosed by bowel biopsy, grade III or IV (4 to 12 years old, both sexes) and 23 healthy children as a control group. A logistic regression model was applied to evaluate an individual's belonging to one group or another. The performance of the model was evaluated by the value of area under the ROC curve. The salivary variables included in the model were the concentration of total proteins, calcium, Ca / P molar ratio, buffer capacity and salivary flow. Results: The total proteins ($p = 0.0016$) and Ca / P molar ratio ($p = 0.0237$) variables were significantly associated with the celiac condition. The value of the area under the ROC curve, estimated from the probabilities of the logistic model, showed that salivary component values allow the celiac condition of patients to be predicted with 85% accuracy ($p < 0.0001$). Conclusion: Logistic discriminant analysis built with salivary variables shows that these are good for predicting this eating pathology with 85% accuracy.

Key words: *Logistic discrimination analysis, diagnosis, celiac disease.*

Introduction

Innovations in molecular biology and statistical models applied to clinical investigations have accelerated the need for more sophisticated anthropological models for assessing biodiversity in relation to human health (1). Multifactorial diseases such as celiac disease (CD), which do not follow Mendelian laws of inheritance, where several different genes interact with environmental and socio-cultural factors, present the challenge of discriminating the additive or interactive effects of different genes, the environment and the socio-cultural context in predisposing to disease (1-3).

CD is a permanent multifactorial disorder, which is associated with genetic factors such as antigens HLA-DQ2 or HLA-DQ8 of T lymphocytes and environmental factors such as the cereal proteins, prolamins (4). Epidemiological data show that CD is a common disease in the world, affecting not only European countries but also people who have European ancestors and live in developing countries like South Asia, South Africa and South America, where their prevalence is similar to the European countries (2). In Argentina, a single screening study in the University of La Plata found prevalence to be 1:160 individuals (5).

Genetic predisposition to this disease has been studied in family clusters, which are found to have a risk 20 to 30 times higher than rest of population. Its early identifica-

tion enables gluten-free diet treatment to begin, reducing the risk of developing lymphoma or gastrointestinal carcinoma (4). An accurate diagnosis is made through bowel biopsy, as other tests, such as genetic or immunological, do not diagnose conclusively (4). Nowadays, researchers are trying to establish non-invasive screening methods to support biopsy indications (6), because the latter is a problematic methodology, particularly in children. In clinical and epidemiological research, statistical methods are also commonly used to establish diagnostic and prognostic criteria for identifying groups with disease (3, 7, 8).

Among biological components, saliva is known to be helpful for the diagnosis of different diseases, and its use has increased remarkably in the last 15 years (9). It is also known to play a protective role in the oral cavity, and factors such as salivary flow as well as organic and inorganic factors are altered in some systemic diseases like CD (4). Antibacterial factors and antibodies have been recognized among the organic components of saliva (10). Al-Bayaty et al. (11) reported the presence of Ig A antigliadin antibodies (IgA AGA) in saliva, suggesting it could be used as an early non-invasive screening test for the detection of the risk of celiac disease in children and adults, to be subsequently confirmed by bowel biopsy.

The hypothesis of the present study was the generation of classification rule, through logistic regression analysis, will help clinicians to estimate the probability of an individual being included in the celiac category, thus avoiding the use of harmful surgical techniques in non-celiac individuals and/or improving the quality of life of CD patients by early diagnosis. The aim was therefore to evaluate the use of some salivary factors as variables for constructing statistical models that can predict the celiac condition of symptomatic patients with a non-invasive methodology.

Materials and Methods

The study group consisted of 52 children with celiac disease (both sexes, 4 to 12 years old) with bowel biopsy diagnosis grade III or IV according to 1990 modified ESPGAN criteria, (12, 13) and 23 children healthy were included as a control group. Both groups attendet at Gastroenterology Service of the Santísima Trinidad Children's Hospital, Córdoba, Argentina, and the Pediatric Dentistry department, Dentistry School, Córdoba National University, Argentina, during the years 2004 and 2005. The healthy controls were children not affected by gastrointestinal disorders or other diseases that can interact with CD and whose serological antigliadin IgG (Ig AGA) and antiendomysium (IgG EMA) tests were negative. We excluded individuals who were medicated with immunosuppressive drugs.

The values of salivary variables were obtained from saliva stimulated with sugar-free chewing gum (Beldent, Stani,®, xylitol mint flavour) for 5 minutes, two hours after meals, by salivation into a graduated polyethylene tube, and kept on ice until processed. Buffer capacity was determined as the difference between pH values before and after the addition of 1ml of 5 mM HCl to 1 ml of saliva (14). To quantify total protein, calcium and phosphate, protocols described by Lowry et al., Ray et al. and Chen et al., respectively, were followed (15-17).

This study was approved by the Research and Ethics Committee of the Córdoba Provincial Health Ministry, in accordance with the Declarations of Nüremberg, Helsinki, and Tokyo of the World Medical Association.

Statistical Model

A binary variable Y (outcome) in this study identified the celiac condition (Y = 1). The predictors variables x (covariables) studied in saliva were total protein concentration (g/L), phosphate (mg%), Ca/P molar ratio, calcium concentration (mg%) and salivary flow expressed as total volume of saliva, in millilitres per minute (ml/min).

The logistic regression model was applied as a discrimination method for assessing the inclusion of individuals in the celiac group. The logistic used in the context of discriminant analysis is the linearity of the log ratio of conditional densities, which in the particular case of the absence-presence of a disease involves the a posteriori odds of an individual being celiac or not (3). The evaluation of accuracy of diagnosis was made by Area Under ROC curve (AUC). In studies like this, where two levels of a condition are evaluated, it is always necessary to set a threshold that discriminates high risk individuals; this is the context of AUC that can be interpreted in this case as the probability of randomly selecting individuals suffering CD (18). For data analysis, the software used was the SSlogis (stats) R version 2.5.0, 2007 package (www.r-project.org) and Infostat professional version 2007.

Results

Average values of salivary components are shown in Table 1. In general, there is a decrease of protein values and Ca/P molar ratio associated to CD. In order to estimate the probability of the celiac condition, the following Logistic Regression model was built:

$$\text{logit (celiac)}=\beta_o + \beta_1 protein + \beta_2 phosphato + \beta_3 ratioCa/P + \beta_4 calcium + \beta_5 capbuf + \beta_6 salflow$$

The estimation of parameters was made by applying the logit link function, with 8 iterations and without missing data. The estimated values of parameters are shown in Table 2.

The protein (p = 0.0016) and Ca / P molar ratio (p = 0.0237) variables were significantly associated with celiac status (Table 2). In relation to these variables, a child presenting lower values of protein and Ca/P molar ratio in saliva will, therefore, has have a high probability of CD (Figure 1, A and B). The estimated AUC value from

Table 1. Mean values ± standard deviation of salivary factors in celiac and control group.

Salivary Factor	*Celiac*	*Healthy Control*
	Average ± SD	
Protein (g/L)	1.35 ± 0.94	2.60 ± 1.12
Phosphate (mg%)	7.22 ± 4.10	6.95 ± 2.27
Molar Ratio Ca/P	0.61 ± 0.62	0.92 ± 0.81
Calcium (mg%)	3.47 ± 2.97	3.06 ± 1.67
Buffer capacity	0.98 ± 0.70	0.70 ± 0.36
Total Volume (ml/min)	4.08 ± 1.69	3.79 ± 1.72

Table 2. Estimated value of the parameters of the logit model built the corresponding standard deviation, the values of the estimated upper and lower bounds of the confidence interval with 95% confidence by Wald Statistic. The p-value estimated with Wald Statistic (*). Highlighted in bold, the p-values considered significant at 95% confidence.

Parameters	Est.	SD	Wald LB (95%)	Wald UB (95%)	p-value(*)
Protein (g/L)	-1.02	0.32	0.19	0.68	**0.0016**
Phosphate (mg%)	-0.07	0.12	0.74	1.17	0.5514
Ratio Ca/P	-2.20	0.97	0.02	0.75	**0.0237**
Calcium (mg%)	0.38	0.22	0.94	2.26	0.0939
Buffer	1.67	1.16	0.55	51.13	0.1485
Total Volume (ml/min)	0.19	0.21	0.81	1.81	0.3573

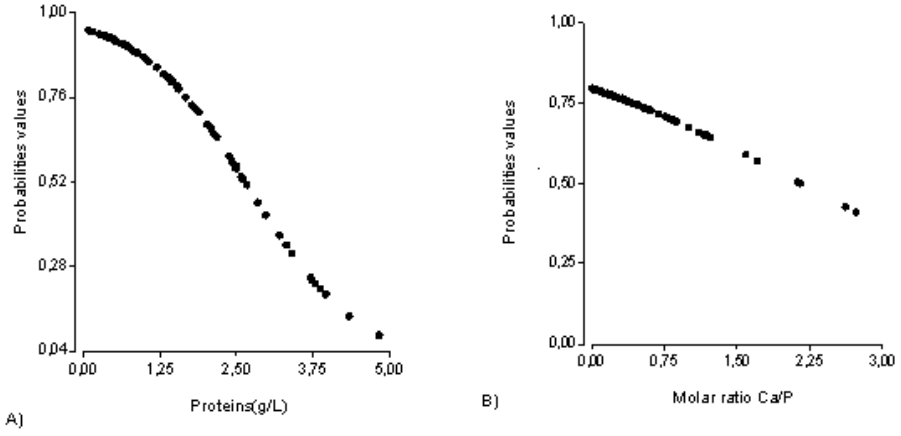
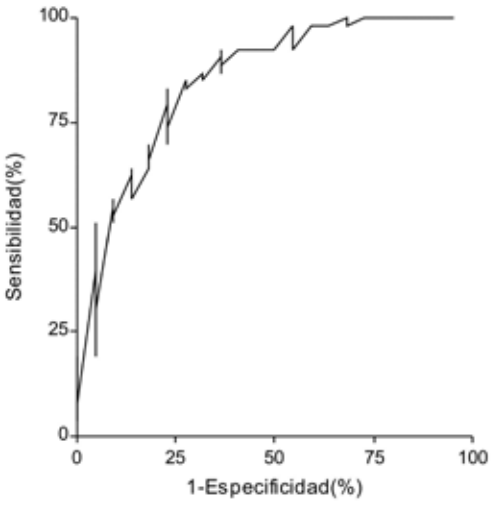


Fig. 1. Curve of probability values calculated by logistic regression model by protein variable (g/L) (A) and molar ratio Ca/P (B).



AUC	p-value
0.8542	**<0.0001**

Fig. 2. Prediction ROC Curve for celiac patients. p-values lower than 0.05 reject Ho: $AUC_{celiac}=AUC_{healthy\ controls}$.

probabilities calculated by the logistic model showed that values of salivary factors enable the celiac condition to be predicted in a patient with 85% accuracy ($p<0.0001$) (Figure 2).

Discussion

CD presents a challenge for classifying individuals on the basis of a classifier, because affected individuals present alterations in several genes that are differentially expressed in a range of phenotypic features and are often seen late, complicating the systemic and buccal health of the patients (19).

Logistic discriminant analysis, built on salivary variables, shows that these are valuable for predicting CD, with a diagnostic accuracy of 85%. However, a 15% classification error represents a significant percentage of individuals who are not properly diagnosed. The logistic discriminant analysis has been created from variables corresponding to a complex genetic disease which may be influencing its performance. However, logistic regression models for predicting probability for a particular condition were and are widely used because they are more flexible and easier to understand and implement than other classification methods (3, 20). Moreover, statistical models offer advantages in relation to clinical predictors as they enable a large number of predictive and explanatory variables of a particular condition to be assessed. However, the use of these tools in the biomedical and dental context is not fully accepted due to ignorance of the benefits that they provide in general and dental medicine (21).

CD being a systemic disease, can impact negatively on various components of the buccal ecosystem. Among the little literature on salivary components, there is a lack of agreement on their value as indicators of the celiac condition (22). In our study, variables such as total protein concentration and Ca/P molar ratio were shown to be associated with the CD condition. In agreement with this, Lenander-Lumikari et al. (22) argue that CD patients who follow a strict gluten-free diet secrete lower levels of amylase, myeloperoxidase, IgA and IgM in stimulated saliva relative to control groups.

In relation to inorganic salivary components, the observed decrease in Ca/P molar ratio values in celiac children compared to healthy controls could be related to the alteration of bone metabolism, a characteristic of this disease (23). It may also be that the reduction in buffer capacity decreases ion bicarbonate concentration, not keeping of salivary pH values. These factors are unfavourable for calcium and phosphorus precipitation. This coincides with the observations of Lenander-Lumikari (22). Our results lead us to conclude that salivary variables are valuable in predicting this eating pathology, diagnosing with 85% accuracy.

References

1. Thornton-Wells TA, Moore JH, Haines JL. Genetics, statistics and human disease: analytical retooling for complexity. Trends Genet. 2004 Dec;20(12):640-7.
2. Cataldo F, Montalto G. Celiac disease in the developing countries: a new and challenging public health problem. World J Gastroenterol. 2007 Apr 21;13(15):2153-9.
3. MacLachlan GJ. Discriminant Analysis and Statistical Pattern Recognition. 1th ed. USA:John Wiley & Sons, Inc; 2004.
4. Dewar D, Pereira SP, Ciclitira PJ. The pathogenesis of coeliac disease. Int J Biochem Cell Biol. 2004 Jan;36(1):17-24.
5. Gomez JC, Selvaggio GS, Viola M, Pizarro B, la Motta G, De Barrio S, et al. Prevalence of celiac disease in Argentina: screening of an adult population in the La Plata area. Am J Gastroenterol. 2001 Sep;96(9):2700-4.
6. Aliaga ED, Miquel BP, Ribes-Koninckx C. Marcadores serológicos de enfermedad celiaca. Acta Pediátrica Española. 2003; 61(1): 24-32.
7. Asparoukhov O. A Comparison of Discriminant Procedures for Binary Variables. Computational Statistics & Data Analysis. 2001; 38:139-60.
8. Marshall RJ. The use of classification and regression trees in clinical epidemiology. J Clin Epidemiol. 2001 Jun;54(6):603-9.
9. Streckfus CF, Bigler LR. Saliva as a diagnostic fluid. Oral Dis. 2002 Mar;8(2):69-76.
10. Dorronsoro S. Ambiente bucal: Equilibrio vs Desequilibrio. Revista Dental Chile. 1997; 88(1): 12-2747.
11. Al-Bayaty HF, Aldred MJ, Walker DM, Newcombe RG, Swift G, Smith PM, et al. Salivary and serum antibodies to gliadin in the diagnosis of celiac disease. J Oral Pathol Med. 1989 Dec;18(10):578-81.
12. [No authors listed] Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. Arch Dis Child. 1990 Aug;65(8):909-11.
13. Perera DR, Weinstein WM, Rubin CE. Symposium on pathology of the gastrointestinal tract-Part II. Small intestinal biopsy. Hum Pathol. 1975 Mar;6(2):157-217.
14. Ericcson Y. Clinical investigations of the salivary buffering action. Acta Odontol Scand. 1959 Sept;17(2):133-65.
15. Lowry OH, Rosebrough NJ, Farr AL, Randall RJ. Protein measure-

ment with the Folin phenol reagent. J Biol Chem. 1951 Nov;193(1):265-75.
16. Chen PS, Toribara TY, Warner H. Microdetermination of phosphorus. Anal Chem. 1959 Nov;28(11):1756-8.
17. Sarkar BC, Chauhan UP. A new method for determining micro quantities of calcium in biological materials. Anal Biochem. 1967 Jul;20(1):155-66.
18. Sullivan Pepe M. The statistical evaluation of medical tests for classification and prediction. 1th ed. New York: Oxford University Press Inc; 2003.
19. Troncone R, Auricchio S. Enfermedad Celíaca, capítulo 25. En: Wyllie R. Gastroenterología Pediátrica. 2th ed. Madrid : McGraw-Hill Interamericana; 2001.
20. Worth AP and Cronin MTD. The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effects. J Mol Struc (Theochem). 2003; 622:97-111.
21. Liao L, Mark DB. Clinical prediction models: are we building better mousetraps. J Am Coll Cardiol. 2003 Sep 3;42(5):851-3.
22. Lenander-Lumikari M, Ihalin R, Lähteenoja H. Changes in whole saliva in patients with coeliac disease. Arch Oral Biol. 2000 May;45(5):347-54.
23. Green PH, Jabri B. Coeliac disease. Lancet. 2003 Aug 2;362(9381):383-91.

Acknowledgments
This work was supported by the Secretaría de Ciencia y Técnica de la Universidad de Córdoba, Argentina (Res SECYT 162/06 – Res Rect 2245/06).