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 University of Cordoba-Argentina
(2) Regular Professor. Buccal Biology Department. Faculty of Dentistry. National University 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 University of Cordoba -Argentina

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-
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 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, have a high probability of CD (Figure 1, A and B). The estimated AUC value from 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 SStologis (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:

\[ \text{logit (celiac)} = \beta_0 + \beta_1\text{protein} + \beta_2\text{phosphato} + \beta_3\text{ratioCa} / P + \beta_4\text{calcium} + \beta_5\text{capbuf} + \beta_6\text{salflow} \]

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, have a high probability of CD (Figure 1, A and B). The estimated AUC value from

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 attend at Gastroenterology Service of the Santisima 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 1 ml 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 SStologis (stats) R version 2.5.0, 2007 package (www.r-project. org) and Infostat professional version 2007.
Table 1. Mean values ± standard deviation of salivary factors in celiac and control group.

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

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.

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

Fig. 1. Curve of probability values calculated by logistic regression model by protein variable (g/L) (A) and molar ratio Ca/P (B).
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 its performance. However, logistic regression models for predicting probability for a particular condition were

![AUC p-value](image)

Fig. 2. Prediction ROC Curve for celiac patients. p-values lower than 0.05 reject Ho: AUC_celiac=AUC_healthy controls.

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
15. Lowry OH, Rosebrough NJ, Farr AL, Randall R.J. Protein measure-

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).