# Advancing Salivary Biomarker Development with Logistic Regression – Machine Learning

Qi Yan[1], Craig S. Miller, D.M.D., M.S.[2], Xiaohua Douglas Zhang, Ph.D[1].

[1]Department of Biostatistics, [2]Division of Oral Medicine, University of Kentucky, Lexington, KY

## Introduction

- Periodontitis is a chronic inflammatory disease that affects ~50% of adults resulting in inflammation, bleeding, and connective tissue destruction.
- Traditional methods of diagnosing periodontitis often involve invasive procedures such as using dental instruments to measure pocket depths.
- Saliva testing eliminates the need for invasive procedures, making it more comfortable for patients and reducing the risk of infection.
- Risk calculators are based on large population data that utilize modifiable and non-modifiable factors to help determine the level of risk for disease presence, progression, and complications.

## Objective

- To determine the best combination of salivary biomarkers that can be used to create a risk calculator for detecting periodontitis non-invasively.

## Methodology

- Prospective multicenter cohort study.
- Demographic, clinical, radiographic findings and whole unstimulated saliva were collected from 90 adult patients ( > age 30) at UK College of Dentistry.
- Samples were assayed for relevant biomarkers (12 protein, 14 bacteria) by Luminex and PCR.
  - Smokers were excluded.
- Logistic regression analysis of different combinations of biomarkers along with demographic and clinical risk factors.
  - Validated with 1,000 iteration bootstrap simulation
  - Average accuracy, sensitivity, and specificity obtained along with confusion matrix
  - Biomarker importance/weight determined

## Results

- A set of saliva biomarkers including one protein and two bacteria, with the joint consideration of age, BMI, and number of teeth as risk factors, achieved a sensitivity of 0.96, specificity of 0.98, and overall accuracy of 0.97 (Table 1 and Table 2).
- The two selected bacteria have the highest contribution to the performance of the logistic model (Figure 1)
- Including risk factors into the model increased sensitivity by ~8% for all 3 combinations
- The best logistic model (based on sensitivity) is:

$$\log(\frac{p}{1-p})=2.92+0.053*Age+0.114*BMI-0.081*\#Teeth+626.7*Bacteroide\ 1$$
$$+1680*Bacteroide\ 2 +0.00003*Protein\ 1$$

Table 1. Results showing the sensitivity, specificity and accuracy of different biomarker combinations

| Risk factors | Biomarker 1 | Biomarker 2 | Biomarker 3 | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|---|---|
| X | Protein 1* | X | X | 0.68 | 0.77 | 0.57 |
| X | Bacteroide 1 | Bacteroide 2 | X | 0.88 | 0.82 | 0.93 |
| X | Bacteroide 1 | Bacteroide 2 | Bacteroide 3 | 0.89 | 0.86 | 0.93 |
| Age + BMI + #Teeth | Protein 2** | X | X | 0.86 | 0.84 | 0.87 |
| Age + BMI + #Teeth | Protein 2 | Bacteroide 1 | X | 0.93 | 0.9 | 0.95 |
| Age + BMI + #Teeth | Protein 2 | Bacteroide 1 | Bacteroide 2 | 0.97 | 0.95 | 0.98 |

X in the cell mean column was not a part of the model. Specific names of biomarkers are omitted as research is under patent. Collaborating company requested biomarkers be limited to a maximum of three.

\* connective tissue remodeling protein inhibitor

\*\* connective tissue remodeling protein

Table 2. Average confusion matrix for the best combination after 1,000 iteration bootstrap

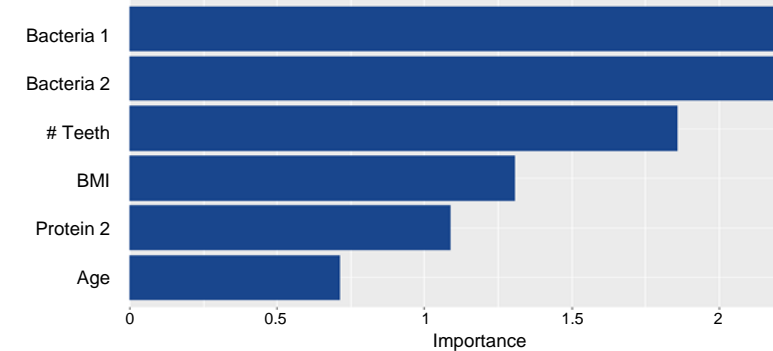| Observed | Predicted | |
|---|---|---|
| | Periodontitis | Healthy |
| Periodontitis | 26.07 | 1.91 |
| Healthy | .92 | 27.1 |



Figure 1. Variable importance of the logistic model for the best combination

## Conclusion and Discussion

- Identified a set of saliva biomarkers with a high sensitivity and specificity that can be used for diagnosing periodontitis non-invasively
- Adding risk factors and/or more biomarkers increased sensitivity

Considerations & Limitations

- Financial costs of biomarker testing taken into consideration.
  - Fewer number of biomarkers needed vs. increased sensitivity
- The A1C (risk factor) was not collected in healthy patients

Future directions

  - Gingivitis discrimination
  - Use machine learning methods such as Random Forest, LASSO.

## Acknowledgements

## References

Hung M, Factors Associated with Periodontitis in Younger Individuals: A Scoping Review. Journal of Clinical Medicine. 2023; 12(20):6442. https://doi.org/10.3390/jcm12206442