U.PORTO
FACULDADE DE
MEDICINA DENTÁRIA
UNIVERSIDADE DO PORTO

Artigo de investigação

Cotinine: Exploring the impact of smoking habits on periodontal disease

Cotinina: Exploração do impacto dos hábitos tabágicos na doença periodontal

João Pedro Teixeira Alves Onofre

Mestrado integrado em Medicina Dentária

Faculdade de Medicina Dentária da Universidade do Porto

U.PORTO
FACULDADE DE MEDICINA DENTÁRIA
UNIVERSIDADE DO PORTO

Mestrado integrado em Medicina Dentária

Cotinine: Exploring the impact of smoking habits on periodontal disease

Autor

Nome completo: João Pedro Teixeira Alves Onofre

Orientador

Nome: José António Ferreira Lobo Pereira

Grau académico: Doutorado em Medicina Dentária pela Faculdade de Medicina Dentária da Universidade do Porto

Título profissional: Professor Auxiliar da Faculdade de Medicina Dentária da Universidade do Porto

Área científica: Periodontologia

Porto, 2024

Agradecimentos

Gostaria de começar por expressar a minha profunda gratidão ao Senhor Professor José António Lobo Pereira, cuja motivação inabalável foi fundamental para que eu pudesse enfrentar um projeto de investigação como este. Agradeço a sua dedicação incansável, as inúmeras horas passadas a trabalhar juntos, a paciência para esclarecer as minhas dúvidas e a constante disponibilidade e dedicação ao longo do último ano e meio. Carrego comigo não só os seus ensinamentos, mas também a amizade e companheirismo que floresceram ao longo do curso.

Pai, Mãe, vocês foram a minha principal fonte de motivação desde o primeiro dia em que cheguei à FMDUP. Um simples “obrigado” não é suficiente para expressar toda a minha gratidão pelos vossos esforços que me permitiram chegar até aqui. Estes cinco anos árduos, longe de vocês, culminaram na realização do meu maior sonho: tornar-me um profissional na área da saúde. Vocês foram, são e sempre serão os meus heróis. Agradeço por terem incutido em mim os valores de humildade, foco, determinação e trabalho, que foram cruciais nesta longa caminhada. Amo-vos incondicionalmente.

Marta, tornaste esta aventura mais colorida. Não imaginas o quanto foste importante para mim nesta jornada. É com o coração cheio que nos vejo concluir esta etapa da vida juntos. Foste o meu porto seguro, confidente de todas as horas, namorada e, sem dúvida, a melhor surpresa que a faculdade me reservou. És um ser humano sem igual. Não posso deixar de agradecer também à Gabriela, ao Rui, à Cristina, à Joana e à Inês, por me fazerem sentir em casa.

Guilherme, meu "irmão", desde o primeiro dia de faculdade até que a vida nos separe, não tenho palavras para expressar o quanto gosto de ti. Tornaste-te o melhor amigo e irmão que nunca tive. Aos meus “pais adotivos”, Amélia e Eduardo, sou eternamente grato por me terem acolhido nos tempos de confinamento. Foram dois anos inesquecíveis e cruciais para que eu não desistisse do sonho. Sem vocês, não seria possível concluir este caminho. Eternamente grato, podem contar comigo para tudo o que precisarem.

Ao Élvio, à Bia, à Francisca e à Andreia, o meu profundo agradecimento pela sincera amizade. Levo-vos comigo para a vida; foram incansáveis.

À família do Krav Maga, o meu sincero obrigado por me tornarem mais forte, por aliviarem os finais de dia e por me motivarem a ultrapassar os meus limites. Ana, Pedro, os responsáveis por terem tornado a minha pessoa mais disciplinada. À Catarina, Bernardo, Helena, Ana Cláudia, Filipa, Bessa, Gonçalo "Paranhos", Monteiro, Andreia, Judo, juntos rumo à faixa preta.

Não poderia deixar de agradecer à Tânia e ao Márcio, por terem acreditado em mim para ajudar na Livingtours.

RESUMO

Introdução: O tabagismo é o principal fator de risco comportamental para o início e agravamento da doença periodontal. Desde o século XX, a investigação tem demonstrado o impacto significativo do tabaco no início, progressão e gravidade dos distúrbios periodontais. A cotinina, o principal metabolito da nicotina, serve como um biomarcador para a exposição ao tabaco devido à sua semi-vida mais prolongada em comparação com a nicotina, sendo mensurável no sangue, saliva ou urina. A periodontite é uma doença inflamatória crónica que destrói progressivamente os tecidos periodontais, constituídos pela gengiva, cemento, ligamento periodontal e osso alveolar.

Objetivo: Este estudo teve como objetivo determinar se a profundidade de sondagem das bolsas periodontais (PPDmax) e a perda de inserção clínica (CALmax) de um indivíduo estão associadas aos níveis séricos de cotinina.

Metodologia: Utilizando dados do inquérito NHANES 2011-2012, empregámos o modelo GAMLSS para analisar o comportamento das variáveis dependentes profundidade de sondagem e da perda de inserção em relação aos níveis séricos de cotinina, género, idade e hábitos tabágicos. Para modelos que não eram previsíveis com GAMLSS, explorámos abordagens de modelagem alternativas e selecionámos o modelo CART. O CART é adequado para captar relações não lineares e interações entre variáveis, fornecendo uma indicação clara da importância das variáveis e permitindo uma interpretação direta.

Resultados: A cotinina (β = 3.30e-09; valor-p = 4.57e-09) mostrou um efeito positivo estatisticamente significativo na profundidade de sondagem máxima (PPDmax). Comparando os dois modelos GAMLSS para profundidade de sondagem máxima, o modelo com idade como fator apresentou um AIC inferior (3715.331) em comparação com o modelo com hábitos tabágicos como variável (AIC = 3744.620), sugerindo que a idade é um preditor mais significativo para PPDmax do que os hábitos tabágicos quando se considera a complexidade do modelo.

No modelo GAMLSS para a perda de inserção clínica máxima (CALmax), os níveis de cotinina (β = 3.2e-03; valor-p = 1.72e-13) sugeriram uma relação positiva entre os níveis de cotinina e CALmax, indicando que níveis mais altos de cotinina estão associados a um aumento da perda de inserção clínica. O género feminino (β = -0.967; valor-p < 2e-16) foi associado a um CALmax inferior em comparação com os homens, sugerindo diferenças de género na saúde periodontal. A idade (β = 0.0541; valor-p < 2e-06) indicou que o CALmax aumenta com a idade, implicando que indivíduos mais velhos tendem a ter maior perda de inserção clínica. A análise dos hábitos tabágicos (β = 2.6e-03; valor-p = 1.48e-07) indicou uma relação positiva significativa entre o tabagismo e CALmax.

Conclusão: Os resultados sugerem que a cotinina está significativamente associada ao aumento da severidade da doença periodontal, evidenciada por valores mais altos de PSmax e CALmax. Os modelos GAMLSS confirmaram essa associação, que permaneceu robusta mesmo ao considerar idade, género e hábitos tabágicos. Os modelos CART identificaram ainda a cotinina como o fator mais crítico na influência da severidade da doença periodontal em comparação com outras variáveis. Estes achados destacam o impacto prejudicial do uso de tabaco na saúde periodontal.

ABSTRACT

Introduction: Smoking is the primary behavioural risk factor for the onset and exacerbation of periodontal disease. Since the 20th century, research has demonstrated the significant impact of tobacco on the initiation, progression, and severity of periodontal disorders. Cotinine, the primary metabolite of nicotine, serves as a biomarker for tobacco exposure due to its longer half-life compared to nicotine, being measurable in blood, saliva, or urine. Periodontitis is a chronic inflammatory disease that progressively destroys periodontal tissues, including the gingiva, cementum, periodontal ligament, and alveolar bone.

Objective: This study aimed to determine whether the probing pocket depth (PPDmax) and clinical attachment loss (CALmax) of an individual are associated with serum cotinine levels.

Methodology: Using data from the NHANES 2011-2012 survey, we employed the GAMLSS model to analyze the behavior of the dependent variables probing depth and attachment loss in relation to serum cotinine levels, gender, age, and smoking habits. For models that were not predictable with GAMLSS, we explored alternative modeling approaches and selected the CART model. CART is suitable for capturing nonlinear relationships and interactions between variables, providing a clear indication of variable importance, and allowing straightforward interpretation.

Results: Cotinine (β = 3.30e-09; p-value = 4.57e-09) showed a statistically significant positive effect on maximum probing pocket depth (PPDmax). Comparing the two GAMLSS models for PPDmax, the model with age as a factor presented a lower AIC (3715.331) compared to the model with smoking habits as a variable (AIC = 3744.620), suggesting that age is a more significant predictor for PPDmax than smoking habits when considering model complexity. In the GAMLSS model for maximum clinical attachment loss (CALmax), cotinine levels (β = 3.2e-03; p-value = 1.72e-13) suggested a positive relationship between cotinine levels and CALmax, indicating that higher cotinine levels are associated with increased clinical attachment loss. Female gender (β = -0.967; p-value < 2e-

16) was associated with lower CALmax compared to males, suggesting gender differences in periodontal health. Age (β = 0.0541; p-value < 2e-06) indicated that CALmax increases with age, implying that older individuals tend to have higher clinical attachment loss. The analysis of smoking habits (β = 2.6e-03; p-value = 1.48e-07) indicated a significant positive relationship between smoking and CALmax.

Conclusion: The results suggest that cotinine is significantly associated with increased severity of periodontal disease, evidenced by higher values of PPDmax and CALmax. The GAMLSS models confirmed this association, which remained robust even when considering age, gender, and smoking habits. The CART models further identified cotinine as the most critical factor influencing the severity of periodontal disease compared to other variables. These findings highlight the detrimental impact of tobacco use on periodontal health.

ABREVIATIONS

APCI - Atmospheric Pressure Chemical Ionization

CAL - Clinical Attachment Loss

CALmax - Clinical Attachment Loss Maximum

CALmean - Clinical Attachment Loss Mean

CART - Classification and Regression Trees

CEJ - Cement Enamel Junction

EGF - Epidermal Growth Factor

FGM - Free Gengival Margin

GAMLSS - Generalized Additive Models for Location, Scale, and Shape

HPLC** - High-Performance Liquid Chromatography

ID- Isotope Dilution

ID HPLC-APCI MS/MS - Isotope-Dilution High-Performance Liquid Chromatography Coupled with Atmospheric Pressure Chemical Ionization Tandem Mass Spectrometry

IFN-γ - Interferon-Gamma

IL-1 - Interleukin-1

IL-6 - Interleukin-6

MDA - Malondialdehyde

MMP - Matrix Metalloproteinase

MS - Mass Spectrometry

MS/MS - Tandem Mass Spectrometry

NCHS - National Center for Health Statistics

PDGF - Platelet-Derived Growth Factor

PGE2 - Prostaglandin E2

PPDmax - Probing Pocket Depth Maximum

PPDmean - Probing Pocket Depth Mean

PMNs - Polymorphonuclear Leukocytes

ROS - Reactive Oxygen Species

TNF-α - Tumour Necrosis Factor-Alpha

CONTENTS

I | INTRODUCTION

INTRODUCTION

Smoking is recognized as a risk factor for both the onset and exacerbation of periodontal disease (1,2). According to the World Health Organization, tobacco use presents one of the gravest public health challenges globally, resulting in over 8 million deaths annually. Of these, more than 7 million deaths are directly linked to active smoking, while approximately 1.3 million are due to exposure to second-hand smoke (3).

Research into the relationship between tobacco use and periodontal health began in the 20th century, highlighting tobacco's significant impact on the initiation, progression, and severity of periodontal disorders (4). Smoking is the principal behavioural risk factor for periodontitis, impacting its prevalence, extent, and severity.

Cotinine, an alkaloid found in tobacco, is the primary active metabolite of nicotine, serves as a biomarker to measure tobacco smoke exposure, with a minimal amount being excreted through the kidneys (5,6). Cotinine provides a more precise measure of nicotine intake than self-reported smoking habits, due to its detectability in blood, saliva, or urine (7). The extended half-life of cotinine, ranging from 15-20 hours—contrasted with nicotine's shorter half-life of approximately 2 hours—establishes it as a reliable biochemical marker for assessing nicotine consumption (8,9).

Periodontitis is defined as a "chronic multifactorial inflammatory disease associated with dysbiotic biofilm and characterized by the progressive destruction of the tooth-supporting apparatus" (10). The periodontal apparatus, comprising the gingiva, cementum, periodontal ligament, and alveolar bone, plays a critical role in maintaining dental stability within the oral cavity (4)

The pathogenesis of periodontitis is initiated by gingivitis, a condition characterized by localized gingival inflammation to dental biofilms. Without clinical intervention, this preliminary inflammatory phase can evolute into chronic periodontitis. This advancement is marked by the emergence of profound periodontal pockets, a critical indicator of progressed **periodontal disease**, potentially leading to tooth loss (11). The pathophysiological framework of this

progression is characterized by the **dysbiose** intricate interaction between the pathogenic biofilm and the host's immune response, perpetuating a chronic inflammatory state that amplifies the systemic inflammatory burden. (11)

The study aimed to determine if the depth of an individual’s periodontal pockets and the Clinical Attachment Loss (CAL) are associated with serum cotinine levels, age of smokers, gender, and duration of smoking.

II | PERIODONTAL DISEASE

PERIODONTAL DISEASE

Periodontal Protocol Assessment

The National Health and Nutrition Examination Survey (NHANES) 2011-12 protocol for periodontal assessment, entails a comprehensive full-mouth periodontal evaluation targeting adults aged 30 to 80 and is executed by qualified dental professionals (Doctor of Dental Surgery / Doctor of Medicine in Dentistry (D.D.S./D.M.D.)).

The assessment employs a HU-Friedy periodontal probe in conjunction with a dental mirror to accurately measure periodontal health indices (12).

The examination protocol begins at the most distal tooth in each quadrant and progresses anteriorly towards the midline. Each quadrant is systematically dried, and the periodontal status of each tooth site—distal, mid, and mesial, both facially and lingually—is evaluated sequentially. Measurements are recorded from the free gingival margin (FGM) to the cement-enamel junction (CEJ), and from the FGM to the base of the pocket. Any recession exposing the CEJ is noted with negative values. This collection of periodontal data is crucial for calculating attachment loss using specialized software, enhancing the precision of periodontal disease diagnosis (12).

The NHANES protocol stipulates that examiner round their measurements to the nearest millimetre using a color-coded NIDR (National Institute of Dental Research) periodontal probe, which has markings at 2, 4, 6, 8, 10, and 12 millimetres. This probe is used to measure periodontal pocket depths and gingival recession at predetermined sites on each tooth. The protocol ensures that the probe is kept parallel to the long axis of the tooth to prevent any angular discrepancies that could affect measurement accuracy, especially in interproximal areas (12).

The assessment is conducted systematically across all quadrants of the mouth, starting with the upper right quadrant and progressing in a tooth-by-tooth sequence towards the midline, continuing through the upper left, lower left, and lower right quadrants. This protocol ensures a comprehensive evaluation of periodontal health, building a robust dataset for the analysis of the effects of

smoking, as evidenced by cotinine levels, on periodontal disease outcomes. The methodological of this approach is critical for producing reliable data, which is necessary for examining the complex relationships between smoking habits and periodontal health (12).

Physiopathology of periodontal disease

Periodontal disease impacts approximately 90% of the global population. (4, 13, 14).

Clinically, healthy gums are pink, firm, and exhibit a well-defined contour around gingival margin. The interdental papillae are intact, do not bleed upon gentle probing, and adequately fill the space under the points of interdental contact, with probing depths ranging from 0 to 3 mm. (15)

From a histological viewpoint, healthy gums display no inflammatory signs, achieving this condition in humans requires extensive and supervised professional intervention over several weeks. In routine clinical practice, gums may appear clinically healthy yet demonstrate an inflammatory infiltrate upon histological examination, primarily consisting of neutrophils in the junctional epithelium and lymphocytes in the adjacent connective tissue. (16)
Within 10 to 20 days of bacterial plaque accumulation, most individuals exhibit clinical indicators of gingivitis, although susceptibility varies, with some showing natural resistance and others more prone to develop clinical signs of the condition (17). Gingivitis is marked by gingival redness and swelling, and an increased probability of soft tissue bleeding during gentle probing (applied pressure = 0.25N). These signs are reversible with the effective removal of bacterial plaque and the implementation of appropriate plaque control strategies. (18)

While early clinical indicators might be subtle, histological changes are pronounced. There is a marked increase in the infiltrate of inflammatory cells, predominantly lymphocytes, macrophages, and neutrophils. (15)

Histologically, gingivitis is distinguished from periodontitis mainly by the increased presence of neutrophils. These cells migrate from adjacent blood vessels, traverse the connective tissue, reach the junctional epithelium, and

extend into the gingival sulcus (19). When the junctional epithelium's barrier function is compromised, allowing for bacterial invasion and their by-products, potentially leading to tissue damage, and initiating a chronic inflammatory response. Neutrophils, however, are typically considered to have a protective function. (19)

There are also significant alterations in the vascular components of the tissue and the cell types within the connective tissues. Blood vessels often proliferate and increase in permeability, facilitating fluid exchange between the blood and the gingival connective tissue. (19)

Gingivitis progresses to periodontitis when periodontal sulcus exceeding 3 mm becoming periodontal pockets, which may deepen over time, thus establishing an expanded anaerobic ecological niche (15). At this stage, there is apical migration of the junctional epithelium from the enamel-cementum junction and lateral and apical expansion of the inflammatory infiltrate into the connective tissue (16). The inflammatory infiltrate in periodontitis contains a higher quantity and greater diversity of inflammatory cells compared to gingivitis, with a notable decrease in neutrophils and an increase in macrophages, T and B lymphocytes, plasmocytes (predominantly), and dendritic cells (20). Furthermore, the intensity and chronicity of inflammation in periodontitis are greater, reflecting the severity and progressive nature of the condition, which is irreversible.

Periodontitis Pathogenesis

The pathogenesis of periodontitis involves multiple mechanisms that contribute to the degradation of the extracellular matrix. This leads to the loss of connective tissue and bone resorption. Key mechanisms include enzymatic processes, the action of reactive oxygen species, phagocytosis, and the release of inflammatory mediators. Each plays a specific role in the progression of the disease. (21)

Enzymatic processes are fundamental in the pathogenesis of periodontitis, particularly through the action of matrix metalloproteinases (MMPs). MMPs are enzymes responsible for degrading components of the extracellular matrix. In periodontitis, these enzymes are produced by polymorphonuclear leukocytes

(PMNs), macrophages, and endothelial cells in response to inflammation. For example, PMNs release MMP-8, which degrades type I collagen, a key structural component of the periodontal extracellular matrix (15). Moreover, macrophages not only produce MMPs but also release cytokines that stimulate resident cells to produce additional MMPs, further exacerbating tissue degradation. (19)

The bacterial flora associated with periodontitis includes pathogens such as *Porphyromonas gingivalis, Treponema denticola, Tannerella forsythia, and Aggregatibacter actinomycetemcomitans*. These bacteria constitute the Sokransky red complex, known for their strong association with periodontitis and their ability to form a subgingival biofilm with a higher pathogenic potential. These microorganisms have sophisticated immune evasion mechanisms (15,19,21), such as altering chemotactic signalling peptides and inhibit the phagocytic function of PMNs, allowing them to persist in the periodontal environment, contributing to the destruction of periodontal tissues and manipulating the host's immune responses to ensure their survival. (21)

Reactive oxygen species (ROS) are primarily produced by PMNs and macrophages as part of the immune response. Although ROS play a role in pathogen destruction, excessive ROS can cause collateral tissue damage by oxidizing proteins, lipids, and DNA, contributing to the oxidative stress observed in periodontitis. (16)

Inflammatory mediators, including tumour necrosis factor-alpha (TNF-α), interferon-gamma (IFN-γ), interleukin-1 (IL-1), prostaglandin E2 (PGE2), platelet-derived growth factor (PDGF), and epidermal growth factor (EGF), significantly affect periodontal turnover. TNF-α and IL-1 enhance MMP production and bone resorption. PGE2 exacerbates periodontal inflammation and promotes bone loss, while PDGF and EGF are important for cell proliferation and repair, though their effectiveness may be compromised in a chronically inflamed environment. (15, 19)**.**

The interaction between the host immune response and specific pathogens in periodontitis triggers a complex cascade of events leading to the progressive destruction of the periodontium.

III | COTININE

COTININE

Tobacco is the primary behavioural risk factor for periodontitis. Tobacco and tobacco smoke comprise a complex mixture of over 9,500 chemical compounds. (22)

Cotinine is the main metabolite of nicotine. Approximately 70-80% of nicotine introduced into the human body is converted into cotinine (23). Due to its longer half-life (16-20 hours) compared to nicotine (30 minutes-3 hours), cotinine is extensively utilized as a biomarker for assessing exposure to tobacco smoke (24,25).

Measurements of cotinine concentration in blood, urine, and saliva are useful in population studies aimed at determining the degree of exposure to smoking habits (24,25). Owing to its extended half-life, cotinine levels remain stable throughout the day, unlike nicotine levels, which may vary based on the time of day and the interval since nicotine was last administered (26).

Cotinine results from nicotine metabolization, mainly during the inhalation of cigarette smoke, but it can also result from the consumption of tomatoes, green peppers, and potatoes (27,28,29). Cotinine is produced because of metabolic processes of nicotine that occur mainly in the liver (70-75%). However, in small amounts, nicotine is also bio transformed by the lungs, kidneys, nasal mucosa, and brain 23).

Ideal biomarkers of cigarette

An ideal biomarker to assess the degree of exposure to cigarette smoke should have the following characteristics: moderate half-life in the body, show specificity, and be assessed in body fluids through analytical methods, additionally, the presence of other compounds should not affect the assessment of the biomarker, and at the same time, the biomarker should not be influenced by external factors others than tobacco smoke (5).

Biomarkers used to assess the body's exposure to smoking include: nicotine, carbon monoxide, carboxyhemoglobin, thiocyanates, and nicotine derivatives (cotinine, trans 3'hydroxycotinine) (30).

Cotinine biomarker of choice

Determining cotinine concentration is useful for assessing the exposure to tobacco smoke. This has been demonstrated in studies where the level of exposure to tobacco smoke was determined in both smokers and passive smokers. Cotinine concentration is determined using various biological materials - urine, blood, saliva, semen, and even hair (31).

Elements of biological plausibility between cotinine and periodontal disease

This part will focus in evidence of cotinine biological effects on biological systems of the host responsible for tissular homeostasis. Can be classified in regulators of oxidation mechanisms, pro-inflammatory mediators, and chemokines.

It was demonstrated that elevated levels of cotinine had a significant effect on decreasing (superoxide dismutase) (SOD) (32). This study analysed the biochemical parameters among 200 healthy participants, comprising equal numbers of smokers and non-smokers (n = 100 each), none of whom had hypertension, dyslipidemia, or diabetes mellitus. It focused on the effects of tobacco exposure, indicated by cotinine levels, on oxidative stress and inflammation—factors known to contribute to periodontal disease. Elevated cotinine levels in smokers were correlated with increased production of interleukin-6 (IL-6), a cytokine that exacerbates inflammatory responses, and higher levels of malondialdehyde (MDA), a marker of oxidative stress resulting from lipid peroxidation (32).
Concurrently, the study observed a decrease in SOD activity, an enzyme crucial in mitigating oxidative damage by neutralizing superoxide anions into less reactive molecules, hydrogen peroxide, and oxygen (32).

SOD functions as a critical antioxidant defence, specifically targeting the superoxide anion radical (33). Under normal conditions, SOD converts this radical into hydrogen peroxide and molecular oxygen, effectively reducing the potential for oxidative damage (34). However, in smokers, the increased production of free radicals, as evidenced by elevated cotinine and MDA levels, overwhelms the antioxidant capacity of SOD (35). This imbalance leads to a

reduction in SOD activity, diminishing its protective efficacy and contributing to the inflammatory and degenerative processes observed in periodontal disease.
The mechanistic pathways through which smoking exacerbates periodontal inflammation and tissue degradation, underscoring the importance of smoking cessation in the management and prevention of periodontal disease.
Preliminary evidence suggests that SOD may act as an inhibitory agent in neutrophil-mediated inflammation, where activated neutrophils adhere to the vascular endothelium, transmigrate to the extravascular space, and release ROS, protease enzymes, and a significant quantity of chemokines (36,37).

Cotinine nhames methodology

Cotinine concentrations tend to be higher (3-8x) in urine than in serum; however, for studies requiring a quantitative assessment of exposure, plasma or serum is considered the fluid of choice. Plasma was chosen for cotinine analyses in NHANES, study due to its advantages in (38):

1. Better standardized samples, i.e., ensuring consistency and comparability.
2. More accurate quantification of cotinine concentration
3. Greater specificity than urine, facilitating the detection of lower levels of cotinine and identifying minimal tobacco exposure.

The NHANES 2011-2012 study employed a detailed protocol for measuring serum cotinine using isotope-dilution high-performance liquid chromatography coupled with atmospheric pressure chemical ionization tandem mass spectrometry (ID HPLC-APCI MS/MS) (38).

This method involves several steps (38,39,40,41):

1. Isotope Dilution (ID): A quantitative technique where a sample is spiked with a known quantity of an isotopically labelled version of the analyte. The internal standard, chemically identical to the target analyte but distinguishable by its isotopic composition, aids in correcting variances during sample preparation and analysis, enhancing quantification accuracy.

2. High-Performance Liquid Chromatography (HPLC): A technique for separating mixture components based on their interactions with a stationary phase (column) and a mobile phase (solvent), useful for analysing complex mixtures.
3. Atmospheric Pressure Chemical Ionization (APCI): APCI is an ionization method used in mass spectrometry (MS) where a sample is ionized by chemical reactions at atmospheric pressure. It is particularly effective for less polar molecules. In APCI, the sample solution is introduced into the ion source, where it is nebulized into small droplets. These droplets are exposed to a corona discharge, ionizing the solvent and analytes.
4. Tandem Mass Spectrometry (MS/MS): Ions generated by the ion source are selected based on their mass-to-charge ratio by the first mass analyser, fragmented, and the fragments are analysed by a second mass analyser, providing detailed molecular structure information, enhancing both specificity and sensitivity of the analysis.
5. Procedure:

   - Sample Preparation: Adding the isotopically labelled standard to the analyte mixture.

   - HPLC Separation: Injecting the prepared sample into the HPLC system to separate components, as they pass through the column.

   - Ionization: The eluted compounds from the HPLC are then ionized using APCI, which makes them detectable by mass spectrometry.

   - Mass Spectrometric Detection: The ionized molecules are detected and analysed in the tandem mass spectrometer. The first mass analyser selects the ion of interest, which is then fragmented, with the fragments analysed by the second mass spectrometer.

   - Quantification: Calculating the ratio of the intensity of the signal from the target analyte to that of the labelled internal standard, yielding an accurate measure of the analyte concentration.

This protocol is highly effective for detecting and quantifying small molecules in complex biological, environmental, or chemical matrices with high precision,

making it invaluable in research and diagnostic applications where accuracy is critical.

IV | STUDY METHODOLOGY

STUDY METHODOLOGY

Data source

The data explored in this study were obtained from the National Health and Nutrition Examination Survey Data for the years 2011-2012, accessed through the National Center for Health Statistics (NCHS) in Hyattsville, MD, from the U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

The inclusion criteria for the study were adults aged between 30 and 80 years who were smokers. Only edentulous individuals were excluded from the study.

Analysis software

For statistical analysis, the R software was employed, which is well-recognized in the field of statistics and data analysis. R is a programming language specifically designed for data analysis and statistics, with built-in capabilities for a wide range of statistical analyses. Its extensive collection of packages extends its capabilities, allowing for the creation of high-quality graphics and integration with other programming languages and platforms. Additionally, its flexibility and customization make it suitable for analysing large datasets and facilitate the reproducibility and documentation of analyses, applicable across various fields for different types of data analysis.

In this study, the R software (R Core Team 2023) packages were used for statistical modelling, such as dplyr (Wickham et al. 2023) for efficient organization and manipulation of datasets, semi-parametric models for location scale and shape (GAMLSS) (Rigby and Stasinopoulos 2005) for statistical modelling and Classification and Regression Trees (CART) models, a non- parametric machine learning that enable to Classification and Regression Trees, ggplot2 (Wickham 2016) and lattice for creating graphics and data visualization, nhanesA (Endres 2023) for accessing and analysing data from the NHANES, and splines (R Core Team 2023). for implementing regression models that include spline functions.

Computational ressources

The primary system utilized for processing and analysis consists of a machine powered by an Intel(R) Core(TM) i9-9900K CPU, operating at a base frequency of 3.60GHz. This processor is well-suited for handling complex computational tasks, including data processing and simulation required in this research.

Accompanying the processor, the system is equipped with 32.0 GB of RAM, ensuring efficient multitasking and smooth execution of memory-intensive applications. The workstation runs on a 64-bit Windows 11 Pro operating system, version 21H2, providing a stable and secure platform for all computational activities. This setup not only supports the demanding requirements of advanced analytical software but also ensures that the computations are carried out with optimal speed and accuracy, crucial for the successful completion of this study.

Statistical method

The GAMLSS framework is employed to analyse the data.

In this study, we utilized a comprehensive set of variables to analyze periodontal health outcomes, specifically focusing on measures such as probing pocket depth maximum (PPDmax), probing pocket depth mean (PPDmean), clinical attachment loss maximum (CALmax), and clinical attachment loss maximum (CALmean). The variables used included cotinine levels, age, gender, and smoking habits, alongside the periodontal variables derived from the NHANES dataset.

V | RESULTS

RESULTS

Descriptive analysis

Table 1: Gender Proportions and Proportion Test Results

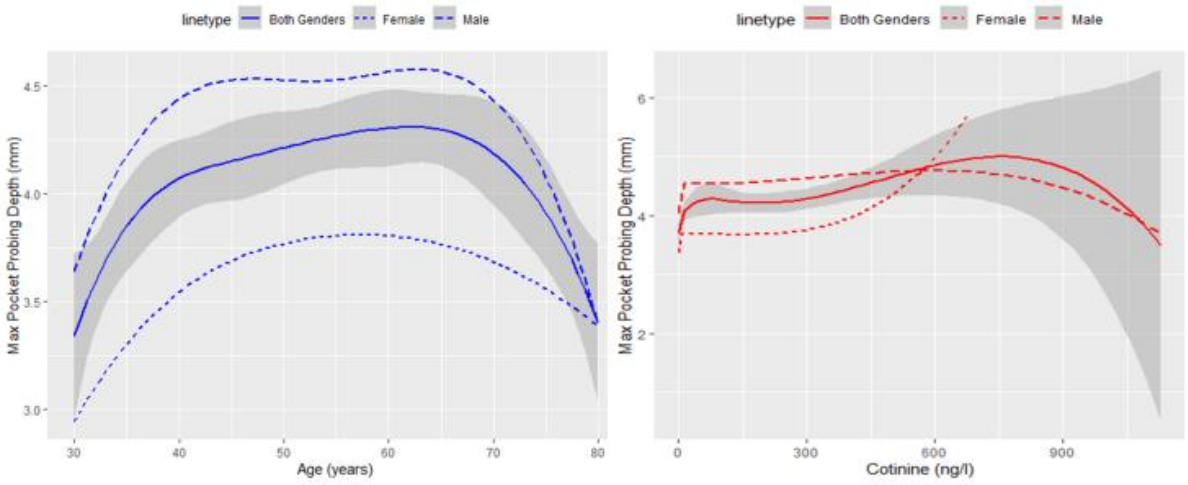
	Count	Proportion
Male	835	0.624
Female	504	0.376

Proportion Test Results	
Statistic	Value
χ^2	81.329
df	1
p-value	$< 2.2e-16$
95% CI for p	(0.597, 0.650)
Sample Proportion (p)	0.624

Table 2: Descriptive Statistics for Study Variables

Var.	Mean	SD	Med.	Trim.	MAD	Min	Max	Skew	Kurt.	SE
Age	52.68	14.15	52.50	52.20	17.05	30.00	80.00	0.18	-0.96	0.39
PPDm	1.65	0.67	1.50	1.57	0.55	0.39	5.41	1.25	2.22	0.02
PPDM	4.06	1.62	4.00	3.90	1.48	1.00	12.00	1.07	1.28	0.04
CALm	1.99	1.22	1.64	1.78	0.74	0.43	11.42	2.42	8.58	0.02
CALM	4.91	2.35	4	4.59	1.48	1	17	1.34	2	0.04
Cot.	108.14	154.20	0.78	79.82	1.14	0.01	1130.00	1.50	2.38	4.22
Dura	34.90	14.91	34.00	34.42	16.31	0.00	80.00	0.23	-0.63	0.41

Graph 1: Graphs of Exploratory Analysis of PPDmax in Relation to Cotinine Levels and Age



PPDmax: pb (Cotinine) + as.factor(Gender) + pb(Smok), Inverse Gaussian model

Table 3: PPD max pb (Cotinine) + as.factor(Gender) + pb(Smok), Inverse Gaussian model

Coeff.	Est.	SE	t-value	$Pr(>\|t\|)$	CI 95%
Mu (Link function: log)					
Intercept	1.429	0.016	88.484	$< 2e-16$	[0.715, 2.144]
Cotinine	$3.30e-04$	$5.59e-05$	5.901	$4.57e-09$	$[1.65e-04, 4.95e-04]$
Female	-0.186	0.021	-8.944	$< 2e-16$	$[-0.227, -0.145]$
Smoke	$3.01e-05$	$9.06e-05$	0.332	0.740	$[-1.51e-05, 7.53e-05]$
Sigma (Link function: log)					
Intercept	-1.619	$2.29e-2$	-70.793	$< 2e-16$	$[-1.664, -1.574]$
Cotinine	$-4.95e-04$	$1.09e-04$	-4.555	$5.72e-06$	$[-6.93e-04, -2.97e-04]$
Model Fit Statistics: No. of observations: 1339; Degrees of Freedom for the fit: 12; Residual Degrees of Freedom: 1327; Global Deviance: 4652.759; AIC: 4676.46; SBC: 4738.08.					

Note: CI 95%: 95% Confidence Interval; SE: Standard Error; Est.: Estimate; Coeff.: Coefficient.

Table 4: Summary of Model Evaluation Metrics

Metric	Average	Standard Deviation	Range
RMSE	1.574	0.067	[1.482, 1.653]
MAE	1.218	0.031	[1.176, 1.257]
MAPE	34.024	0.731	[33.301, 34.816]
R^2	0.046	0.037	[0.005, 0.105]
AIC	3744.620	16.199	[3723.814, 3763.478]

Note: RMSE – Root Mean Square Error; MAE – Mean Absolute Error; MAPE – Mean Absolute Percentage Error; R^2 – Coefficient of Determination; AIC – Akaike Information Criterion.

PPDmax: pb (Cotinine) + as.factor(Gender) + pb(Age), Inverse Gaussian model

Table 5: GAMLSS Model Summary for PPDmax

Coeff.	Est.	SE	t-value	$Pr(>\|t\|)$	CI 95%
Mu (Link function: log)					
Intercept	1.326	0.042	31.734	$< 2e-16$	$[1.193, 1.459]$
Cotinine	$3.02e-04$	$5.58e-05$	5.412	$7.38e-08$	$[1.510e-04, 4.529e-04]$
Female	-0.185	0.020	-9.015	$< 2e-16$	$[-0.092, -0.277]$
Age	$1.97e-03$	$7.10e-04$	2.777	$5.56e-03$	$[9.855e-04, 2.957e-03]$
Sigma (Link function: log)					
Intercept	-1.626	0.023	-71.272	$< 2e-16$	$[-0.813, -2.440]$
Cotinine	$-5.44e-04$	$1.09e-04$	-4.972	$7.48e-07$	$[-2.719e-04, -8.157e-04]$

Model Fit Statistics: No. of observations: 1339; Degrees of Freedom for the fit: 11.68; Residual Degrees of Freedom: 1327.32; Global Deviance: 4618.962; AIC: 4642.316; SBC: 4703.031.

Note: CI 95%: 95% Confidence Interval; SE: Standard Error; Est.: Estimate; Coeff.: Coefficient.

Table 6: Cross- Validation Test

Metric	Average	Standard Deviation	Range
RMSE	1.543	0.064	[1.476, 1.628]
MAE	1.190	0.038	[1.143, 1.241]
MAPE	33.351	1.072	[31.724, 34.486]
R^2	0.084	0.036	[0.057, 0.145]
AIC	3715.331	17.428	[3695.742, 3735.064]

Note: RMSE – Root Mean Square Error; MAE – Mean Absolute Error; MAPE – Mean Absolute Percentage Error; R^2 – Coefficient of Determination; AIC – Akaike Information Criterion.

PDmean: Age

Table 7: GAMLSS Model Summary for PPDmean

Coeff.	Est.	SE	t-value	$Pr(>\|t\|)$	CI 95%
Mu (Link function: log)					
Intercept	0.254	0.044	5.720	$1.32e-08$	[0.127, 0.381]
log(Cotinine)	0.0436	$3.9e-03$	11.211	$<2e-16$	[0.022, 0.065]
Female	−0.201	0.020	−10.010	$<2e-16$	[-0.101, -0.302]
Age	$2.64e-03$	$7.2e-04$	3.659	**2.63e-04**	[0.001, 0.004]
Sigma (Link function: log)					
Intercept	−1.002	0.027	−37.400	$<2e-16$	[-0.501, -1.502]
log(Cotinine)	−0.0137	0.0075	−1.827	0.0679	[-0.007, -0.021]

Model Fit Statistics: No. of observations: 1339; Degrees of Freedom for the fit: 8.25; Residual Degrees of Freedom: 1330.75; Global Deviance: 2165.752; AIC: 2182.252; SBC: 2225.149.

Note: CI 95%: 95% Confidence Interval; SE: Standard Error; Est.: Estimate; Coeff.: Coefficient.

Table 8: Summary of Recent Model Evaluation Metrics

Metric	Average	Standard Deviation	Range
RMSE	0.630	0.040	0.573 to 0.668
MAE	0.455	0.015	0.431, 0.469
MAPE	29.111	1.049	[27.845, 30.302]
R^2	0.122	0.034	[0.081, 0.165]
AIC	1745.871	8.941	[1736.250, 1757.259]

Note: RMSE – Root Mean Square Error; MAE – Mean Absolute Error; MAPE – Mean Absolute Percentage Error; R^2 – Coefficient of Determination; AIC – Akaike Information Criterion.

PPDmean: Smok

Table 9: Summary of Model CART

Métrica	Valor
Número Total de Observações	1.339
Método	ANOVA
Parâmetro de Complexidade (CP)	Inicial: 0,066, Final: 0,010
Número de Divisões	3
Erro Relativo	Inicial: 1,000, Final: 0,875
Erro Cruzado Validado (xerror)	Inicial: 1,001, Final: 0,889
Erro Padrão de xerror (xstd)	Inicial: 0,056, Final: 0,054

Table 10: Importance of Variables

Variável	Importância (%)
LBXCOT	47
as.factor(RIAGENDR)	42
SmokDura	10

Table 11: Detailed Analysis of Nodes

Nó	Observações	CP	Média de PPDmean	MSE	Divisão Primária	Melhoria
1	1.339	0,066	1,651117	0,452	Cotinina $< 0,1785$	0,066
2	612	0,021	1,462645	0,353	Sexo como RL	0,060
3	727	0,038	1,809775	0,481	Sexo como RL	0,066
4	238	N/A	1,280	0,273	N/A	N/A
5	374	N/A	1,579	0,369	N/A	N/A
6	266	N/A	1,575921	0,304	N/A	N/A
7	461	N/A	1,944711	0,534	N/A	N/A

Table 12: Cross- Validation Test

Metric	Average	Standard Deviation	Range
RMSE	0.632	0.079	0.577 to 0.767
MAE	0.477	0.042	0.445 to 0.550
MAPE	32.489	1.716	30.682 to 34.635

CALmax: Age

Table 13: GAMLSS Model Summary for CALmax

Coeff.	Est.	SE	t-value	$Pr(>\|t\|)$	CI 95%
Mu (Link function: identity)					
Intercept	2.153	0.200	10.762	$< 2e-16$	[1.077, 3.230]
pb(Cotinine)	$3.2e-03$	$4.0e-04$	7.447	$1.72e-13$	[0.003, 0.005]
as.factor(Gender)2	-0.967	0.100	-9.716	$< 2e-16$	[-0.630, -1.451]
pb(Age)	0.0541	0.0036	14.884	$< 2e-16$	[0.029, 0.081]
Sigma (Link function: log)					
Intercept	-0.925	0.021	-44.876	$< 2e-16$	[-0.509, -1.387]
pb(Cotinine)	$-2.e-04$	$9.0e-05$	-1.747	0.081	[-8.310e-05, -2.351e-04]
Nu (Link function: identity)					
Intercept	-0.138	0.068	-2.039	0.042	[-0.074, -0.207]
pb(Cotinine)	$1.1e-03$	$3.5e-04$	3.226	$1.29e-03$	[5.882e-04, 1.696e-03]
Tau (Link function: log)					
Intercept	0.881	0.070	12.59	$< 2e-16$	[0.477, 1.321]
Model Fit Statistics: No. of observations: 1339; Degrees of Freedom for the fit: 14.28121; Residual Degrees of Freedom: 1324.719; Global Deviance: 5529.733; AIC: 5558.295; SBC: 5632.553.					
Note: CI 95% columns are not provided in the R output and thus are represented as '–'. SE: Standard Error; Est.: Estimate; Coeff.: Coefficient. Bolded $Pr(>\|t\|)$ values indicate significance below the 0.05 level, keeping our statistical spirits high and our p-values low.					

Table 14: Summary of model Assessment Metrics

Metric	Average	Standard Deviation	Range
RMSE	2.277	0.152	2.067 to 2.480
MAE	1.700	0.086	1.607 to 1.791
MAPE	33.662%	1.395	31.353% to 34.902%
R^2	0.178	0.039	0.121 to 0.232
AIC	4455.444	11.103	4444.436 to 4469.764

CALmax: Smok

Table 15: GAMLSS Model Summary for CALmax

Coeff.	Est.	SE	t-value	$Pr(>\|t\|)$	CI 95%
Mu (Link function: identity)					
Intercept	4.726	$8.9e-02$	52.576	$< 2e-16$	[2.649, 7.089]
pb(Cotinine)	$3.2e-03$	$5.0e-04$	7.154	$1.40e-12$	[1.744e-03, 4.842e-03]
as.factor(Gender)2	-0.979	0.101	-9.694	$< 2e-16$	[-0.519, -1.469]
pb(Smok)	$2.6e-03$	$5.0e-04$	5.283	$1.48e-07$	[1.858e-03, 3.859e-03]
Sigma (Link function: log)					
Intercept	-0.9189	$2.06e-02$	-44.596	$< 2e-16$	[-0.589, -1.378]
pb(Cotinine)	$-.00e-05$	$9.0e-05$	-0.685	0.494	[-3.951e-05, -9.096e-05]
Nu (Link function: identity)					
Intercept	-0.1816	0.067	-2.728	$6.46e-03$	[-0.122, -0.272]
pb(Cotinine)	$1.2e-03$	$3.0e-04$	3.403	$6.9e-04$	[6.242e-04, 1.8e-03]
Tau (Link function: log)					
Intercept	0.886	0.070	12.71	$< 2e-16$	[0.550, 1.329]
Model Fit Statistics: No. of observations: 1339; Degrees of Freedom for the fit: 17.16385; Residual Degrees of Freedom: 1321.836; Global Deviance: 5576.423; AIC: 5610.751; SBC: 5699.997.					
Note: CI 95%: 95% Confidence Interval. SE: Standard Error; Est.: Estimate; Coeff.: Coefficient. Significant $Pr(>\|t\|)$ values are highlighted.					

Table 16: Cross- Validation Test

Metric	Average	Standard Deviation	Range
RMSE	2.518	0.306	2.254 to 3.028
MAE	1.765	0.079	1.627 to 1.829
MAPE	34.641%	2.452	31.355% to 38.246%
R^2	-0.034	0.346	-0.648 to 0.162
AIC	4514.013	22.828	4499.586 to 4554.017

CALmean

Table 17: Summary of CART Model 2 Statistics

CP	nsplit	rel error	xerror
0.07639165	0	1.0000000	1.0015287
0.06698000	1	0.9236083	0.9618003
0.02274963	2	0.8566284	0.8888876
0.01513126	3	0.8338787	0.8564012
0.01465415	4	0.8187475	0.8532790
0.01329958	5	0.8040933	0.8438628
0.01142415	6	0.7907937	0.8451701
0.01000000	7	0.7793696	0.8360988

Table 18: Variable Importance Partitioning Analysis

Variable	Importance
Cotinina	46%
Age	37%
Género	18%

Table 19: Summary of Recursive Partitioning Analysis

CP	**nsplit**	**rel error**	**xerror**	**xstd**
7.639e-2	0	1.000	1.002	7.004e-2
6.698e-2	1	0.924	0.959	6.454e-2
2.275e-2	2	0.857	0.903	6.095e-2
1.513e-2	3	0.834	0.895	6.034e-2
1.465e-2	4	0.819	0.885	6.013e-2
1.330e-2	5	0.804	0.881	5.998e-2
1.142e-2	6	0.791	0.883	6.038e-2
1.000e-2	7	0.779	0.873	5.974e-2

VI | DISCUSSION

DISCUSSION

The initial step in understanding the population under study involves an exploratory analysis of the demographic and clinical characteristics of the sample. Table 1 presents the gender proportions and the results of the proportion test. The sample consists of 835 males (62.4%) and 504 females (37.6%), indicating a significant gender imbalance ($\chi^2 = 81.329$, df = 1, $p < 2.2e\text{-}16$). The 95% confidence interval for the proportion of males in the sample is (0.597, 0.650), with a sample proportion (p) of 0.624.

Table 2 provides descriptive statistics for key variables, offering insights into the central tendencies measures and distributions within the sample. The mean age is 52.68 years, with a standard deviation of 14.15 years, and a median age of 52.50 years, indicating a middle-aged population with a broad age range (30 to 80 years). The periodontal variables, PPDmean, PPDmax, CALmean, and CALmax, reflect the severity of periodontal disease in the population. PPDmean has a mean of 1.65 mm and a maximum value of 5.41 mm, while PPDmax has a mean of 4.06 mm, with a maximum of 12.00 mm. These measurements suggest variability in periodontal probing depths among participants.

The exploratory analysis indicates that both age and cotinine levels significantly influence the depth of periodontal pockets, with notable differences between genders. PPDmax tends to increase with age until a peak around 55 years, followed by a decline. Similarly, PPDmax increases with cotinine levels up to about 600 ng/l, after which it decreases. In both age and cotinine level analyses, males generally show higher PPDmax values compared to females. These findings suggest that age and tobacco exposure, as measured by cotinine levels, are important factors in periodontal health, with distinct impacts observed between genders.

The clinical attachment loss variables, CALmean and CALmax, exhibit means of 1.99 mm and 4.91 mm, respectively, with significant skewness and kurtosis, particularly for CALmean (skew = 2.42, kurtosis = 8.58). This indicates a wide range of attachment loss severity within the sample, with few participants experiencing severe periodontal degradation.

Cotinine levels show substantial variation with a mean of 108.14 ng/mL and a high standard deviation of 154.20 ng/mL. The median cotinine level is 0.78 ng/mL, indicating that while the majority of participants have low to moderate exposure, a subset exhibits very high cotinine levels (up to 1130.00 ng/mL). The duration of smoking (Dura) has a mean of 34.90 years, further emphasizing the long-term tobacco exposure in this population.

This exploratory analysis highlights the diverse demographic and clinical characteristics of the sample, providing a foundation for understanding the relationships between these variables and periodontal health outcomes. The gender imbalance, age distribution, and significant variation in periodontal and cotinine measures underscore the importance of considering these factors in subsequent analyses and interpretations.

The variables chosen for our research study: PPDmax; PPDmean; CALmax; CALmean; Age; Gender; Smoking habits, all have biological and clinical significance.

The clinical variables mentioned, such as PPDmax and PPDmean, are essential for understanding a patient's periodontal status and possess considerable clinical and biological relevance.

PPDmax indicates the maximum probing depth found at a single location in the mouth. This measurement is critical because an elevated, that surpass a critical value in any part of the mouth may indicate the presence of periodontal pocket, which is a sign of periodontal disease.

Clinically, a PPDmax over 3mm value suggests periodontal pathology and probably require dedicated treatment at that specific location, even if other areas are less affected. However, it does not provide information about the overall condition of the mouth, as it is a point-specific measure.

PPDmean represents the average probing depths throughout the mouth. It is a more conservative measure and provides a general overview of the periodontal status. A PPDmean ≥3 mm is generally considered indicative of poor periodontal health. Elevated mean values suggest a periodontal condition that requires comprehensive clinical management.

CAL, or clinical attachment level and represents the status extent of periodontal loss support around a tooth, indicating the distance from the CEJ to the epithelial attachment. Like PPD, it is essential for diagnosing and monitoring the progression of periodontitis.

CALmax measures the maximum clinical attachment loss found at a single location in the mouth. It is a critical measure to quantify the loss of supporting tissue.

CALmean values above zero throughout the mouth indicate significant loss of supporting tissue and an unfavourable periodontal condition.

PPDmax/mean and CALmax/CALmean are fundamental for establishing an accurate diagnosis and for effective periodontal treatment planning. Regular evaluation of these variables helps monitor disease progression and the effectiveness of therapeutic interventions. It is important that each assessment is conducted considering the patient's history, complete clinical evaluations, and, when necessary, supplementary examinations to obtain a more comprehensive view of the patient's periodontal status.

This research began by analysing the behaviour of the dependent variable PPDmax in relation to serum cotinine levels, gender, and smoking habits through the development of a GAMLSS model, as shown in Table 3.

GAMLSS modelling offers significant advantages, including:

1. Flexibility in Distribution Choice: GAMLSS allows for a wide range of distributions, providing flexibility in modelling different types of data.
2. Accommodates Different Types of Responses: GAMLSS is suitable for various response types, including binary, count, continuous, and survival data.
3. Ability to Model Location, Scale, and Shape Parameters: It provides a comprehensive approach by modelling not just the mean (location) but also the variance (scale) and shape (skewness and kurtosis) of the distribution.

However, when using GAMLSS statistical modelling, we must face the following challenges:

1. Computational Intensity: Fitting GAMLSS models, especially with complex data and multiple parameters, can be computationally demanding.
2. Model Complexity: The complexity of the model can increase rapidly with the addition of parameters for scale and shape, making it harder to interpret.
3. Parameter Selection and Overfitting: Choosing the correct distribution and parameters is crucial and challenging; there is also a risk of overfitting with too many parameters.
4. Requirement for Specialized Software: GAMLSS models often require specialized statistical software and expertise to implement effectively.

Table 3, which presents the GAMLSS Model Summary for PPDmax. In the model, there is a set of three variables: cotinine, age, and gender. This allows for the evaluation of the influence of each variable and the robustness of the models in describing and predicting the PPDmax.

Cotinine (β = 3.30e-09; p-value = 4.57e-09) shows a statistically significant positive effect on PPDmax. This confirms its role as a substantial risk factor for periodontal disease.

Gender (β = -0.186; p-value = 4.57e-09) is indicated to significantly affect PPDmax; being female is associated with lower PPDmax values, suggesting a lower risk of developing severe periodontal disease compared to males. This finding aids in understanding variability in susceptibility to periodontal disease across genders. It could be hypothesized that females might have a protective effect against periodontal disease, potentially leading to a healthier periodontium.
Smoking habits (β = 3.01e-05; p-value = 0.740) did not show a statistically significant effect on PPDmax in this model, as evidenced by the p-value.
This outcome suggests that the direct effects of cotinine levels, may be more pertinent to the periodontal condition than the act of smoking itself.

The statistical rigor in this analysis is evident from the high accuracy achieve for cotinine ((β = 3.30e-09, CI [1.65e − 04, 4.95e − 04]) and for gender (β = -0.186; CI [-0.227, -0.145]), underscoring their significant impact on periodontal disease status.

The analysis of model, demonstrate by table 4 fit statistic shows that a lower global deviance indicates a better fit of the model to the data. The value of 4652.759, while standalone does not indicate goodness unless compared with other models, suggests that the model captures a significant amount of the variability in the data.

AIC is used to evaluate the model's performance, considering the number of parameters used. Lower values generally indicate a better fit considering the complexity of the model. The AIC of 3744.620 suggest that the model is relatively efficient in balancing model complexity with fit. However, without comparison models, these values serve primarily as benchmarks for model refinement.

Table 4 presents the results for the 5-fold cross-validation test of the first GAMLSS model. Cross-validation is a statistical method used to evaluate the generalizability and robustness of a model. It involves partitioning the data into training and testing sets 5 times to ensure that the model's performance is consistent across different subsets of data.

Cross-validation is an essential method for model validation, providing a more accurate estimate of model performance compared to a single train-test split. By evaluating the model on multiple subsets, cross-validation helps in detecting overfitting, ensuring the model performs well on unseen data. It assists in selecting the best hyperparameters for the model, leading to optimal performance. Additionally, cross-validation offers reliable estimates of performance metrics, such as RMSE, MAE, and R^2, by averaging results across different folds, which enhances the stability and trustworthiness of these metrics. This method also facilitates the comparative analysis of different models or algorithms on the same dataset, aiding in the selection of the most appropriate model for the given problem.

The RMSE is defined as the square root of the average squared differences between the observed actual outcomes and the outcomes predicted by the model. Both RMSE and MAE are measures of prediction error, but RMSE is more sensitive to larger errors due to the squaring of differences before averaging. In this context, an RMSE of 1.574 being higher than an MAE of 1.218 indicates that there are some larger errors that have a significant impact on the RMSE value.

The low standard deviation (0.067) indicates that the RMSE values are consistent across different folds, suggesting stability in the model's predictive accuracy. The RMSE values range from 1.482 to 1.653. This range provides insight into the variability of the model's performance across different folds. The relatively narrow range suggests that the model consistently performs within a small margin of error, indicating robustness in its predictive capability.The MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. An average MAE of 1.218 indicates that the model's predictions are, on average, 1.218 units away from the actual values. The small standard deviation (0.031) again suggests that the model's performance is consistent across different folds. The MAE values range from 1.176 to 1.257. This narrow range reflects the consistency and reliability of the model's performance across different validation sets. The MAPE indicates the average percentage error between the predicted and actual values. An average MAPE of 34.024% means that, on average, the model's predictions deviate from the actual values by 34.024%. The MAPE values range from 33.301 to 34.816. This narrow range suggests that the model's relative prediction error is stable across different subsets of the data, indicating robustness in its percentage accuracy. This relatively high percentage suggests that while the model might be reliable in terms of absolute errors, its relative error is considerable, which may be due to the scale of the data or the presence of outliers. The R^2 value indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. An average R^2 of 0.046 suggests that the model explains only about 4.6% of the variance in PPDmax. This low R^2 value indicates that the model has limited explanatory power and that many factors influencing PPDmax are not captured by the model. The R^2 values range from 0.005 to 0.105. This range shows that the model's explanatory power varies across different subsets of the data, from very low to modest levels. The AIC provides a measure of the relative quality of statistical models for a given dataset, with lower values indicating a better-fitting model. The average AIC of 3744.620, with a standard deviation of 16.199, suggests that the model's fit is consistent across different folds, though it is not exceptionally low, indicating there might be room for model improvement. The AIC values range from 3723.814 to 3763.478. This range shows the variability in

the model fit across different subsets of the data, with lower values indicating better model performance.

The cross-validation results validate the current model's consistency but also highlight its limited explanatory power and significant relative error. These findings underscore the need for further model refinement and possibly the inclusion of additional relevant variables to improve predictive accuracy and explanatory capability. However, the primary objective of this study is to determine whether cotinine has a significant effect on selected periodontal disease parameters, rather than to explain the overall variance of PPD in the referred periodontal parameters.

In statistical modelling, the selection of variables is crucial to capturing the relevant factors that influence the dependent variable. In this study, the objective was to analyse the influence of different factors on PPDmax using the GAMLSS model.

Initially, we chose smoking habit as an explanatory variable due to its well-documented relationship with periodontal health. Smoking is a known risk factor for periodontal diseases, and it was expected to significantly influence cotinine levels. Therefore, the inclusion of smoking habit aimed to capture the direct contribution of smoking on PPDmax through the mediation of cotinine.

The initial model, presented in Table 3, smoking habit as variable. The results indicated that smoking habit did not show statistical significance ($p = 0.740$), suggesting a weak or non-existent direct relationship with PPDmax.

Given the low impact of smoking habit observed in the first model, we re-evaluated the model construction and explored the influence of other potentially relevant variables.

The second model, presented in Table 5, replaced smoking habit with age, while retaining cotinine and gender as explanatory variables. The results showed that cotinine and female gender continued to have significant effects on PPDmax and age also demonstrated a significant relationship with PPDmax (β =1.97e-09; p-value= 5.56e-09), indicating that the inclusion of this variable improves the explanation of the variation in periodontal probing depth.

Comparing the two models (Table 3 and Table 5), we observed that the model with age as a factor presented a lower AIC (3715.331) compared to the model with smoking habits as a variable (AIC = 3744.620), suggesting that age is a more significant predictor for PPDmax than smoking habits when considering model complexity. The model with age also showed lower RMSE (1.543) and MAE (1.190) values compared to the model with smoking habits (RMSE = 1.574, MAE = 1.218), further demonstrating that the model with age provides more accurate predictions. Additionally, the model with age as a predictor had a higher (R^2= 0.084) compared to the other model (R^2 = 0.046). The higher R^2 value for the age model indicates that it explains a greater proportion of the variance in PPDmax, highlighting age as a more substantial factor in predicting periodontal probing depth.

The initial choice to include smoking habit was based on existing literature regarding its adverse effects on periodontal health. However, subsequent analysis showed that age is a more relevant and significant predictor in the context of our specific model. Replacing smoking habit with age not only improved the model fit and performance but also provided a more accurate understanding of the factors influencing PPDmax. This approach demonstrates the importance of continuous re-evaluation of statistical models and consideration of different variables to obtain more robust and meaningful insights.

Periodontal probing depth is a critical measure in periodontal health assessment. PPDmax refers to the maximum probing depth recorded in a patient's periodontal examination, while PPDmean represents the average probing depth across all sites examined. Both metrics provide valuable insights, with PPDmax indicating the worst-affected sites and PPDmean offering an overall assessment of periodontal health. The analysis of PPDmean provides insights into the overall periodontal health status, allowing for a more balanced view of disease severity across all measured sites.

Table 7 presents the GAMLSS model summary for PPDmean. The model includes age, gender, and cotinine levels as explanatory variables.

There is a significant positive association between cotinine levels and PPDmean (β = 0.044; p-value = <2e-06). This indicates that higher cotinine levels, reflecting greater exposure to tobacco, are associated with higher average probing depths.

Being female is associated with a significant reduction in PPDmean (β = -0.201; p-value = <2e-06), suggesting gender differences in periodontal health outcomes.

Age shows a significant positive effect on PPDmean (β = 2.64e-03; p-value = 2.63e-04), indicating that average probing depth increases with age, which is consistent with the progression of periodontal disease over time.

When analysing the impact of smoking habits on periodontal health, it is essential to select a statistical model that accurately captures the relationships between variables and provides reliable predictions. The initial attempt to use the GAMLSS model revealed limitations in its predictability when incorporating the smoking habits variable. Consequently, we opted for the CART model. This section justifies this decision and interprets the results within the context of the study.

Given the limitations of the GAMLSS model, we explored alternative modelling approaches and selected the CART because CART is well-suited for capturing non-linear relationships and interactions between variables. Smoking habits may interact with other factors in complex ways that are not easily modelled by GAMLSS. CART provides a clear indication of variable importance, allowing for a more straightforward interpretation of the influence of smoking habits relative to other variables.

The CART model analysis demonstrates that the primary split is based on cotinine levels. This is evident from Table 11, where Node 1, with all 1,339 observations, splits based on cotinine < 0.1785, resulting in an improvement of 0.066. This underscores the significance of cotinine in predicting PPDmean, aligning with findings from the GAMLSS model which also highlighted cotinine as a key predictor. The high variable importance score of 47% for LBXCOT (cotinine) in Table 10 further supports this.

The secondary splits in the CART model are based on gender. In Table 11, Node 2 (612 observations) and Node 3 (727 observations) are both split by gender, with improvements of 0.060 and 0.066, respectively. This indicates that gender

differences are crucial in understanding variations in PPDmean. The importance score for gender (as.factor(RIAGENDR)) is 42% as shown in Table 10, confirming its significance in the model.

Smoking duration, while having a lower importance score of 10% in Table 10, is still relevant in the context of periodontal health. Its inclusion in the CART model highlights its role, albeit to a lesser extent than cotinine and gender. The presence of smoking duration in the variable importance list suggests that it contributes to the overall model, capturing additional nuances in the data.

The detailed node analysis in Table 10 reveals significant variations in PPDmean across the nodes, what reflect the heterogeneity in periodontal health outcomes based on cotinine levels, gender, and smoking duration. The CART model's ability to capture these complex interactions and provide a hierarchical structure of variable importance and splits enhances the understanding of the data.

The choice of the CART model over the GAMLSS model for analysing the impact of smoking habits on periodontal health is justified by CART's capacity to handle non-linear relationships and interactions more effectively. The CART model provides a clearer interpretation of variable importance and reveals critical insights into the influence of cotinine, gender, and smoking duration on PPDmean.

The CART model's structure, as detailed in Table 9, demonstrates significant reductions in error metrics. These improvements in error metrics, along with the detailed node analysis and variable importance scores, provide a robust and meaningful analysis. This approach addresses the limitations observed with the GAMLSS model, offering a more comprehensive understanding of the factors influencing periodontal probing depths.

Comparing the cross-validation metrics, the GAMLSS model (Table 8) shows slightly better and more consistent performance across RMSE, MAE, and MAPE compared to CART model (Table 12). The GAMLSS model also provides the AIC metric, which aids in evaluating model fit with penalization for complexity. The standard deviations of the metrics indicate that the first model exhibits more stable performance across different cross-validation folds. The absence of R^2 and AIC metrics for the second model limits a complete comparison, but the available

metrics suggest the first model is preferable based on lower error rates and greater consistency.

The analysis focuses on the GAMLSS model for CALmax (Table 13), incorporating three explanatory variables: cotinine levels, gender, and age. This model aims to evaluate the impact of these predictors on the maximum clinical attachment loss.

The cotinine levels (β = 3.2e-03; p-value = 1.72e-13) suggests a positive relationship between cotinine levels and CALmax, with a highly significant p-value. This indicates that higher cotinine levels, are associated with increased clinical attachment loss.

The female gender (β = -0.967; p-value = <2e-16) indicates that being female is associated with a lower CALmax compared to males, with high statistical significance. This suggests gender differences in periodontal health, where males tend to exhibit greater clinical attachment loss than females.
The age (β = 0.0541; p-value = <2e-06) indicates that CALmax increases with age, with a significant effect. This implies that older individuals tend to have higher clinical attachment loss.

The last analysis focuses on the GAMLSS model for CALmax, using cotinine levels, gender, and smoking status as the three explanatory variables, instead of age.

Smoking habits (β =2.6e-0.3; p-value = 1.48e-07) indicates a significant positive relationship between smoking and CALmax. This suggests that smoking contributes to increased clinical attachment loss.

Based on the comparison of RMSE, MAE, MAPE, R^2, and AIC, the model that includes age as an independent variable (Table 14) demonstrates better predictive performance and explanatory power for CALmax compared to the model that includes smoking habits (Table 16). Therefore, the model with age is considered the superior model for predicting CALmax in this context.

The CART (Classification and Regression Trees) model provides a non-linear approach to understanding the factors influencing CALmean by recursively partitioning the data into homogenous subsets based on the predictor variables.

The following analysis details the model's performance, variable importance, and recursive partitioning analysis.

The initial relative error is 1.000 (Table 17), which decreases as more splits are made, indicating the model's improved fit with each additional split. The cross-validated error (xerror) follows a similar trend, showing the model's performance stability during validation.
Cotinine (46%) is the most influential variable, followed by age (37%) and gender (18%), table18. This highlights the significant role of tobacco and demographic factors in predicting CALmean.

The table 19 shows that as the model complexity increases (more splits), the relative error decreases, indicating better fit. The cross-validated error (xerror) also decreases, suggesting that the model generalizes well to unseen data. The standard error of the cross-validated error (xstd) remains relatively stable, indicating consistent model performance.

The CART model's recursive partitioning highlights how these variables interact to influence CALmean. For instance, higher cotinine levels combined with older age may result in significantly higher CALmean, identifying high-risk groups that require more intensive periodontal care.

VII | CONCLUSION

CONCLUSION

The results suggest that cotinine is associated with the severity of periodontal disease, being linked to increased values of PPDmax and CALmax .The GAMLSS models demonstrated a significant relationship between cotinine and both PPDmax and CALmax, which remained robust even after the inclusion of additional predictors such as age, gender, and smoking habits. Furthermore, the CART models indicated that cotinine is the most important factor influencing the severity of periodontal disease compared to age, gender, and smoking habits.

These findings can help patients understand the detrimental effects of tobacco use on periodontal health.

VIII | REFERENCES

REFERENCES

1. Grossi S.G., Zambon J., Machtei E.E., Schifferle R., Andreana S., Genco R.J., Cummins D., Harrap G. Effects of smoking and smoking cessation on healing after mechanical periodontal therapy. J. Am. Dent. Assoc. 1997;128:599–607. doi: 10.14219/jada.archive.1997.0259.
2. Grossi S.G., Genco R.J., Machtet E.E., Ho A.W., Koch G., Dunford R., Zambon J.J., Hausmann E. Assessment of Risk for Periodontal Disease. II. Risk Indicators for Alveolar Bone Loss. J. Periodontol. 1995;66:23–29. doi: 10.1902/jop.1995.66.1.23
3. Global Burden of Disease [database.Washington, DC: Institute of Health Metrics; 2019. IHME,accessed 17 november 2023
4. Borojevic T. Smoking and periodontal disease. Mater Sociomed. 2012;24(4):274-6. doi: 10.5455/msm.2012.24.274-276. PMID: 23678331; PMCID: PMC3633395.
5. Benowitz NL, Hukkanen J, Jacob P III. Nicotine chemistry, metabolism, kinetics and biomarkers. Handb Exp Pharmacol. 2009;(192):29-60. doi: 10.1007/978-3-540-69248-5_2. PMCID: PMC2953858. NIHMSID: NIHMS235126. PMID: 19184645.
6. Benowitz NL, Jacob P, Sachs DP. Deficient C-oxidation of nicotine. Clin Pharmacol Ther. 1995;57:590–4
7. Pérez-Stable EJ, Marín G, Marín BV, Benowitz NL. Misclassification of smoking status by self-reported cigarette consumption. Am Rev Respir Dis. 1992 Jan;145(1):53-7. doi: 10.1164/ajrccm/145.1.53. PMID: 1731599
8. Walter C, Kaye EK, Dietrich T. Active and passive smoking: assessment issues in periodontal research. Periodontol. 2000;2012(58):84-92.
9. Eramo S, Tassi C, Negri P, Manta M, Fraschini M, Pedetta F. ELISA analysis of salivary cotinine in smokers. Minerva Stomatol.2000;49:163-168
10. Tonetti MS, Greenwell H, Kornman KS. Staging and grading of periodontitis: Framework and proposal of a new classification and case definition. J Clin Periodontol. 2018 Jun;45 Suppl 20:S149-S161. doi: 10.1111/jcpe.12945. Erratum in: J Clin Periodontol. 2019 Jul;46(7):787. PMID: 29926495

11. Kinane DF, Stathopoulou PG, Papapanou PN. Periodontal diseases. Nat Rev Dis Primers. 2017 Jun 22;3:17038. doi: 10.1038/nrdp.2017.38. PMID: 28805207
12. Centers for Disease Control and Prevention (US). Oral Health Examiners Manual: NHANES 2011-2012 [Internet]. Atlanta: National Center for Health Statistics (US); 2012 [cited 2024 March 15]. Available from: https://wwwn.cdc.gov/nchs/data/nhanes/2011-2012/manuals/Oral_Health_Examiners_Manual.pdf
13. Pihlstrom BL, Tabak L. The National Institute of Dental and Craniofacial Research: research for the practicing dentist. J Am Dent Assoc. 2005;136(6):728-737
14. Listgarten MA. Pathogenesis of periodontitis. J Clin Periodontol. 1986;13(5):418-25
15. Lindhe J, Karring T, Lang NP. Clinical Periodontology and Implant Dentistry. 4th ed. Chapter 5: Interactions Between Parasite and Host in Periodontal Disease. In: Kinane DF, Berglundh T, Lindhe J, editors. 1997. p. 148-175.
16. Page RC, Schroeder HE. Pathogenesis of inflammatory periodontal disease. A summary of current work. Lab Invest. 1976 Mar;34(3):235-49. PMID: 765622
17. Van der Weijden GA, Timmerman MF, Danser MM, Nijboer A, Saxton CA, Van der Velden U. Effect of pre-experimental maintenance care duration on the development of gingivitis in a partial mouth experimental gingivitis model. J Periodontal Res. 1994 May;29(3):168-73. doi: 10.1111/j.1600-0765.1994.tb01209.x. PMID: 8207626.
18. Lindhe J, Rylander H. Experimental gingivitis in young dogs. Scand J Dent Res. 1975 Nov;83(6):314-26. doi: 10.1111/j.1600-0722.1975.tb00444.x. PMID: 1105761.
19. Preshaw PM. Carranza's Clinical Periodontology. 12th ed. Chapter 5: Periodontal Pathogenesis. In: Newman MG, Takei HH, Klokkevold PR, Carranza FA, editors. Philadelphia: Elsevier; 2015.
20. Garant PR, Mulvihill JE. The fine structure of gingivitis in the Bengal. III. Plasma Cell Infiltration of the Subepithelial Connective Tissue. J Periodontol. 1970 Jan;41(1):32-38.

21. Bartold PM, Narayanan AS. Diseased Periodontium. In: Bartold PM, Narayanan AS, editors. Biology of the Periodontal Connective Tissues. Carol Stream: Quintessence Publishing Co. Inc; 1998.
22. Li Y, Hecht SS. Carcinogenic components of tobacco and tobacco smoke: a 2022 update. Food Chem Toxicol. 2022 Jul;165:113179. doi: 10.1016/j.fct.2022.113179. PMCID: PMC9616535. PMID: 35643228.
23. Hukkanen J, Jacob P 3rd, Benowitz NL. Metabolism and disposition kinetics of nicotine. Pharmacol Rev. 2005;57(1):79-115.
24. Benowitz NL. Cotinine as a biomarker of environmental tobacco smoke exposure. Epidemiol Rev. 1996;18:188-204
25. Benowitz NL. Pharmacology of nicotine: addiction, smoking-induced disease, and therapeutics. Annu Rev Pharmacol Toxicol. 2009;49:57-71. doi: 10.1146/annurev.pharmtox.48.113006.094742. PMCID: PMC2946180. PMID: 18834313.
26. Bao Z, He XY, Ding X, Prabhu S, Hong JY. Metabolism of nicotine and cotinine by human cytochrome P450 2A13. Drug Metab Dispos. 2005 Jan;33(1):258-61. doi: 10.1124/dmd.104.002105. PMID: 15528319
27. Yildiz D. Nicotine, its metabolism, and an overview of its biological effects. Toxicon. 2004 Apr 15;43(6):619-32. doi: 10.1016/j.toxicon.2004.01.017. PMID: 15109883
28. Tutka P, Mosiewicz J, Wielosz M. Pharmacokinetics and metabolism of nicotine. Pol J Pharmacol. 2005 Mar-Apr;57(2):143-53. PMID: 15886412
29. Hukkanen J, Jacob P 3rd, Benowitz NL. Metabolism and disposition kinetics of nicotine. Pharmacol Rev. 2005 Mar;57(1):79-115. doi: 10.1124/pr.57.1.3. PMID: 15734728
30. Dhar P. Measuring tobacco smoke exposure: quantifying nicotine/cotinine concentration in biological samples by colorimetry, chromatography and immunoassay methods. J Pharm Biomed Anal. 2004 Apr 15;35(1):155-68. doi: 10.1016/j.jpba.2004.01.009. PMID: 15030890.
31. Vine MF, Hulka BS, Margolin BH, Truong YK, Hu PC, Schramm MM, Griffith JD, McCann M, Everson RB. Cotinine concentrations in semen, urine, and blood of smokers and nonsmokers. Am J Public Health. 1993 Sep;83(9):1335-8. doi: 10.2105/ajph.83.9.1335. PMID: 8363014; PMCID: PMC1694994.

32. Kumboyono, Chomsy IN, Hakim AK, Sujuti H, Hariyanti T, Srihardyastutie A, Wihastuti TA. Detection of vascular inflammation and oxidative stress by cotinine in smokers: measured through interleukin-6 and superoxide dismutase. Int J Gen Med. 2022;15:7319-28. doi: 10.2147/IJGM.S367125. PMCID: PMC9489220. PMID: 36147199.
33. Hussain T, Al-Attas OS, Alamery S, Ahmed M, Odeibat HAM, Alrokayan S. The plant flavonoid, fisetin alleviates cigarette smoke-induced oxidative stress, and inflammation in Wistar rat lungs. J Food Biochem. 2019;43(8):e12962. doi: 10.1111/jfbc.12962
34. Fukai T, Ushio-Fukai M. Superoxide dismutases: role in redox signaling, vascular function, and diseases. Antioxid Redox Signal. 2011;15(6):1583–1606. doi: 10.1089/ars.2011.3999
35. Pawar VS, Sontakke A, Pawar AA, Kakade S. Assessment of serum cotinine and oxidative stress markers in tobacco users. Biomed Pharmacol J. 2019;12:579–586. doi: 10.13005/bpj/1677
36. Lawrence LA, Mulligan JK, Roach C, et al. Superoxide dismutase reduces the inflammatory response to Aspergillus and Alternaria in human sinonasal epithelial cells derived from patients with chronic rhinosinusitis. Am J Rhinol Allergy. 2015;29(2):89–93. doi: 10.2500/ajra.2015.29.4155
37. Yasui K, Baba A. Therapeutic potential of superoxide dismutase (SOD) for resolution of inflammation. Inflamm Res. 2006;55(9):359–363. doi: 10.1007/s00011-006-5195-y
38. Centers for Disease Control and Prevention. Cotinine [Internet]. Atlanta: CDC; 2012 [cited 2024 Jan 15]. Available from: https://wwwn.cdc.gov/nchs/data/nhanes/2011-2012/labmethods/cot_g_met_cotinine.pdf
39. Jin S, Pang W, Zhao L, Zhao Z, Mei S. Review of HPLC-MS methods for the analysis of nicotine and its active metabolite cotinine in various biological matrices. Biomed Chromatogr. 2022 Jun;36(6):e5351. doi: 10.1002/bmc.5351. Epub 2022 Mar 11. PMID: 35106788.
40. Petersen GO, Leite CE, Chatkin JM, Thiesen FV. Cotinine as a biomarker of tobacco exposure: development of a HPLC method and comparison of matrices. J Sep Sci. 2010 Mar;33(4-5):516-21. doi: 10.1002/jssc.200900575. PMID: 20155742.

41. Hawks R. Analytical methodology. En: National Institute on Drug Abuse Monograph Series No. 73. Rockville, MD: National Institute on Drug Abuse; 1986. p. 30-42