

How confounding factors distort clinical phenotyping in Sjögren's syndrome

Kristina Lacasta

Universidad de Sevilla

Angela Gandara-Alvarez

Universidad de Sevilla

Maria J. Rus

Universidad de Sevilla

Cristiane Cantiga-Silva

Universidad de Sevilla

Virginia Moreira Navarrete

Hospital Universitario Virgen Macarena

Carmen Dominguez Quesada

Hospital Universitario Virgen Macarena

Jose Javier Perez Venegas

Hospital Universitario Virgen Macarena

Juan Antonio Ortega

Universidad de Sevilla

Aurea Simon-Soro

`asimon@us.es`

Universidad de Sevilla

Research Article

Keywords: Sjögren's Syndrome, Machine Learning, Confounding Bias, Computational Phenotyping, Explainable AI, Autoimmune Diseases, Aging

Posted Date: October 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11133046/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Importance

Clinical phenotyping in Sjögren's syndrome is complicated by non-specific symptoms that overlap with aging, comorbidities, and medication effects. Understanding how these confounding factors affect diagnostic accuracy is essential for valid observational research.

Objective

To evaluate how age, comorbidities, and xerogenic medications influence clinical phenotyping and diagnostic performance in Sjögren's syndrome.

Methods

Cross-sectional study of 196 women (97 Sjögren's syndrome, 99 controls). Principal component analysis identified latent phenotypic structure. Random forest classifiers with SHAP values assessed variable importance. Multivariable regression estimated independent associations with stimulated salivary flow.

Results

A classifier using all variables achieved AUC 0.92 but relied on diabetes, cardiovascular disease, and medication use. A systemic-only model achieved AUC 0.78. An oral-only model retained AUC 0.82. In regression, Sjögren's diagnosis (-0.52 mL/min; 95% CI -0.73 to -0.29) and xerogenic medication (-0.45 mL/min; 95% CI -0.75 to -0.15) were independently associated with lower stimulated salivary flow; age showed a smaller effect (-0.018 mL/min/year; 95% CI -0.027 to -0.010). Diagnostic accuracy declined after age 70.

Interpretation

Confounding factors can inflate apparent diagnostic performance in Sjögren's syndrome studies. Researchers should consider control-group heterogeneity and audit models for reliance on systemic variables rather than disease-specific oral pathology.

1 Introduction

Sjögren's syndrome is a chronic autoimmune disease characterized by lymphocytic infiltration of exocrine glands, primarily affecting salivary and lacrimal glands[1]. The disease predominantly affects women in their postmenopausal years, with most cases diagnosed around age 55–60[2, 3]. The hallmark symptoms, xerostomia and hyposalivation, are notoriously non-specific in this demographic[4, 5].

Dry mouth symptoms affect up to 30% of adults over 65 years, but objective salivary hypofunction is less common and primarily attributable to medications, diabetes, and cardiovascular disease rather than aging itself[6, 7]. Commonly prescribed medications list xerostomia as a side effect, with risk increasing

substantially with polypharmacy[8]. This creates a diagnostic challenge in the population where Sjögren's syndrome is most prevalent, distinguishing autoimmune glandular damage from medication-induced or comorbidity-related dysfunction is difficult.

When such confounding factors are unevenly distributed between cases and controls, they can distort the observed relationship between Sjögren's syndrome and its oral manifestations. Prediction models developed from such data may exploit these confounders as diagnostic shortcuts, achieving high apparent accuracy without capturing disease-specific oral pathology[9–11].

The aim of this study was to evaluate how age, comorbidities, and xerogenic medications affect clinical phenotyping in Sjögren's syndrome. Using an observational cohort of 196 women, we sought to: (i) identify latent phenotypic structure, (ii) assess whether diagnostic models rely on systemic confounders rather than oral pathology, and (iii) estimate independent associations of disease status, aging, and medication use with salivary dysfunction.

2 Materials and Methods

2.1 Study Design and Population

This observational cross-sectional study analyzed a total cohort of 196 women recruited from Virgen Macarena University Hospitals, Seville (Spain). The cohort was classified into a study group of 97 patients diagnosed with Sjögren's Syndrome (SS) according to the 2016 ACR/EULAR classification criteria, and a control group of 99 non-Sjögren women. To ensure demographic comparability, inclusion criteria for the control group required participants to be women aged 40 years or older at any menopausal stage. Exclusion criteria for both groups included age under 18 years (under 40 for controls), a history of head and neck radiotherapy or chemotherapy, use of parasympathomimetic drugs, recent antibiotic or antifungal treatment within the past three months, or the use of chlorhexidine-based mouth rinses within the last month. All participants provided written informed consent prior to inclusion.

The cohort of patients with Sjögren's Syndrome was approved by the Ethics Committee of the Virgen Macarena-Virgen del Rocío University Hospitals (CEI 0501-N-22), and the cohort of menopausal women without Sjögren's Syndrome was approved by the Ethics Committee of the Virgen de Valme University Hospital, CEI Sevilla Sur (CEI 1532-N-21). All procedures were conducted in accordance with the Declaration of Helsinki.

2.2 Clinical and Oral Assessment

In addition to oral examinations, comprehensive systemic health data were gathered via structured medical questionnaires. This included the documentation on chronic comorbidities (e.g., diabetes, cardiovascular diseases) and current pharmacological treatments. Notably, the active use of medications with well-documented xerogenic side effects was systematically recorded as a binary variable.

Salivary function was evaluated through the collection of unstimulated (SA) and stimulated (ST) whole saliva samples following standardized sialometry protocols. First, saliva by drooling (SA) was collected in a sterile tube for 5 min. Subsequently, stimulated saliva (ST) was collected by chewing paraffin pellets in another sterile tube for 5 min. All samples were kept at 4°C until processing[12].

The salivary flow was measured by the decantation method in a calibrated test tube. A normal salivary flow rate was established at 0.3 mL/min for SA and 1.5 mL/min for ST[13].

Regarding oral health status, clinical data was recorded using a standardized form. Periodontal status was assessed using the Community Periodontal Index (CPI) across six sextants using a calibrated probe[14, 15]. Dental status was evaluated using the DMFT index (Decayed, Missing, and Filled Teeth) [16]. Additionally, the presence of oral mucosal lesions and prosthetic rehabilitation was documented during the examination.

2.3 Data Preprocessing

Data analysis was performed using the Python 3.10 ecosystem, using the pandas, NumPy, SciPy, matplotlib and scikit-learn libraries[17–22]. No variable exceeded 5% missingness. Remaining missing values were imputed using Multiple Imputation by Chained Equations (MICE) to preserve multivariate relationships, accounting for the complex interdependencies inherent in clinical data[23, 24]. Importantly, the target variable (Sjögren’s diagnosis) contained no missing values and was therefore strictly excluded from being imputed, preventing any data leakage or artificial label generation. Following imputation, variables

with zero variance were removed before modeling. To reduce the effect of outliers in categorical-derived variables, data standardization was performed using RobustScaler, which scales features using statistics robust to outliers by centering on the median and scaling to the interquartile range. Finally, to reduce dimensionality while preserving clinical information, the composite DMFT index was retained, and its individual components were removed. Notably, age and menopausal status were both retained despite collinearity to explicitly disentangle their respective impacts on salivary function.

2.4 Computational Framework

The computational strategy was structured in three distinct phases to identify, quantify, and mitigate confounding bias. A global random seed was initialized for all stochastic procedures to support reproducibility.

First, to explore latent data structures without diagnostic bias, Principal Component Analysis (PCA) was performed. PCA is widely utilized in biomedical informatics to reduce dimensionality and uncover hidden structures in highly correlated clinical datasets[25, 26]. Pearson correlation coefficients were calculated between the original clinical variables and the top principal components (PC1 and PC2) to interpret the dominant axes of variance. This approach aims to identify potential clustering driven by age or metabolic comorbidities, independent of the clinical diagnosis.

Second, diagnostic classification was modeled using a Random Forest Classifier. Tree-based ensemble methods are particularly suited for heterogeneous medical datasets due to their robustness to non-linear relationships, ability to handle mixed data types, and resistance to overfitting in moderately sized cohorts[27, 28]. Following a conservative modeling strategy to prioritize generalizability, the algorithm was instantiated with a set of baseline parameters ($n_estimators = 100$, $criterion = 'gini'$, $max_depth = None$, $min_samples_split = 2$, $min_samples_leaf = 1$, and $bootstrap = True$). This baseline approach was intentionally chosen to minimize the risk of overfitting through excessive hyperparameter manipulation; more complex optimization techniques, such as nested cross-validation for hyperparameter tuning, are deferred for future studies with larger cohorts. All stochastic procedures were initialized with a fixed random seed ($random_state = 42$) to ensure reproducibility.

For model evaluation, a Stratified 5-Fold Cross-Validation scheme was implemented to

preserve class balance across all folds and estimate the Area Under the Receiver Operating Characteristic Curve (AUC). Final model discrimination and classification metrics (e.g., accuracy, sensitivity, and specificity) were computed on the independent 20% held-out test set using the standard 0.5 probability threshold, which is appropriate given the balanced nature of our cohort. Additionally, the theoretical optimal operating point was projected on the Receiver Operating Characteristic (ROC) curve using Youden's J statistic for visual reference. A subsequent age-stratified performance analysis evaluated the classifier's accuracy across specific age subgroups.

To quantify the specific weight of confounding factors, a feature ablation strategy was designed by training three distinct model configurations using identical cross-validation folds[29, 30]. The *comprehensive model* included all clinical, demographic, and comorbidity variables to establish baseline performance. The *systemic-only model* was trained using only demographic and comorbidity data, acting as a negative control. The *biological model* was restricted to oral phenotyping variables after the elimination of systemic confounders. A comprehensive inclusion matrix detailing the exact feature subsets utilized for each of these three models is provided in Supplementary Table S1. Model interpretability was primarily driven by SHapley Additive exPlanations (SHAP) values[31, 32], with Mean Decrease in Impurity (MDI) provided as a complementary metric. This dual approach accounts for the known bias of MDI towards continuous variables or features with high cardinality, ensuring a more robust and transparent ranking of feature importance.

Third, to disentangle age-related decline from disease-intrinsic damage, Multivariate Ordinary Least Squares (OLS) Regression was performed using the statsmodels library[33]. Multivariate regression remains a standard approach in epidemiological and clinical studies for adjusting continuous outcomes against multiple covariates, thereby allowing the estimation of independent associations from confounding factors[34–36].

Stimulated salivary flow was modeled as the dependent variable, with age, group, and medication use as covariates, allowing for the estimation of adjusted coefficients to quantify the floor effect (intercept difference) versus the rate of decline (slope).

Finally, standard descriptive and inferential statistical analyses were conducted to characterize the clinical cohort. Continuous variables were summarized using means and standard deviations (SD) or medians and interquartile ranges (IQR), depending on data distribution evaluated via the Shapiro–Wilk test. Group comparisons for continuous data were performed using the independent-samples *t*-test for normally distributed variables and the Mann–Whitney *U* test otherwise. Categorical variables were summarized as frequencies and percentages and compared using Pearson’s chi-square test or Fisher’s exact test, as appropriate. Odds ratios (OR) with 95% confidence intervals (CI) were calculated to quantify effect sizes for selected binary outcomes. Statistical significance was established at a two-sided *p*-value < 0.05 for all conventional tests.

3 Results

3.1 Clinical Characteristics of the Cohort

3.1.1 Demographics and Comorbidities Profile

The study cohort comprised 196 women (mean age 54.64 years), balanced between patients diagnosed with Sjögren’s Syndrome (SS, *n* = 97, mean age 58.87) and a control group (No-SS, *n* = 99, mean age 50.78) (Fig. 1A).

Analysis of general health revealed a significant imbalance in the comorbidity profile (Fig. 1B). The control group exhibited a markedly higher prevalence of metabolic and cardiovascular conditions compared to the SS group. Specifically, diabetes was present in 41.41% of controls versus only 7.22% of SS patients, and cardiovascular diseases were twice as prevalent in controls (56.56% vs 26.80%). Conversely, SS patients showed a higher mean comorbidity count (29.5 ± 23.2) but lower rates of specific lifestyle risk factors. For instance, smoking prevalence was significantly higher in controls (25.25%) compared to SS patients (9.27%). Alcohol consumption followed a similar trend (52.52% in controls vs 34.02% in SS). These findings indicate that the control group presented a prominent metabolic and lifestyle risk phenotype, diverging from a strictly healthy profile.

3.1.2 Oral Health Status and Salivary Function

Regarding oral status, patients with Sjögren’s Syndrome showed a lower frequency of dental prosthesis use (26.05%) compared to controls (49.49%; OR = 0.359, *p* = 0.001). Although mucosal lesions were slightly more frequent in the SS group (17.39% vs 15.15%), this difference did not reach statistical significance. Similarly, periodontal status assessed via the Community Periodontal Index (CPI) revealed extensive pathology in both groups, but no statistically significant association was found between periodontal disease and Sjögren’s diagnosis (OR = 1.24; *p* = 0.539) (Fig. 1C).

Functional assessment of salivary glands highlighted the diagnostic value of stimulated flow. While unstimulated salivary flow showed high rates of hyposalivation in both groups (65.63% in SS vs 58.16% in controls) with no significant mean differences (0.27 vs 0.29 mL/min), stimulated salivary flow

revealed a drastic functional divergence. Hyposalivation under stimulation was present in 92.78% of SS patients compared to 72.45% of controls, with significantly lower mean flow rates in the disease group (0.65 vs 1.1 mL/min; OR = 4.889, $p = 0.0002$) (Fig. 1D). These results highlight that, while stimulated hyposalivation serves as a robust functional marker, the high background prevalence of non-specific oral pathology in the control group masks direct clinical differentiation.

3.2 Unsupervised Analysis Reveals a Metabolic Phenotype Con-founding the Control Group

Principal Component Analysis (PCA) was utilized as an interpretable exploratory tool to examine the latent structure of the dataset, rather than to establish a definitive diagnostic separation. This unsupervised analysis revealed two dominant confounding axes (Fig. 2A) that, while accounting for a modest proportion of the total variance (14.8% and 8.8% for PC1 and PC2, respectively), effectively capture the primary sources of clinical bias. The first component (PC1) captured a gradient of cumulative oral deterioration, where variables such as the number of sound teeth (loading = 0.84) and the DMFT index (loading = 0.81) showed strong collinearity with age (loading = 0.66) (Fig. 2B). This confirms that biological aging acts as a primary demographic confounder, creating a pervasive overlap between patients and controls.

The second component (PC2) was driven by systemic comorbidities, including diabetes ($r = 0.48$), cardiovascular disease ($r = 0.56$), and hyposalivation-inducing medications ($r = 0.48$) (Fig. 2C). Bootstrap loading stability analysis confirmed robust positioning of age and oral deterioration variables on PC1 (SNR > 18), and stable positioning of cardiovascular disease and diabetes on PC2 (SNR > 3) (Fig. 2D). Variance decomposition by variable group showed that lifestyle variables account for the largest share (37.5%), followed by dental health (25.0%) and comorbidities & medication (20.8%) (Fig. 2E). Consequently, the separation of Sjögren's patients in the latent space was not driven by oral health status alone but by their lack of this specific metabolic and pharmacological burden, exposing a phenomenon of metabolic confounding in the control population.

3.3 Shortcut Learning: Systemic Comorbidities Inflate Diagnostic Performance

The Random Forest classifier trained on the full dataset achieved a mean cross-validated accuracy of 0.78 and an AUC of 0.92, suggesting high apparent diagnostic performance (Fig. 3A). However, inspection of the model's decision structure revealed that this performance was driven primarily by systemic variables (Fig. 3B). Diabetes emerged as the root node of the decision tree, acting as the primary exclusion criterion (Fig. 3D). Feature importance analysis revealed that systemic variables (Diabetes, Hyposalivation-inducing medication, and cardiovascular disease) dominated the predictive hierarchy (Fig. 3C). The confusion matrix on the held-out test set showed key diagnostic metrics (Fig. 3E).

To rigorously measure the magnitude of this confounding effect, we trained a negative control model using only demographic and comorbidity data. This systemic-only model achieved an AUC of 0.78. This confirms that a substantial proportion of the discriminative power in standard datasets can be derived solely from patient history, without examining the oral cavity.

3.4 Validation of a Biological Oral Signature: Performance of the De-confounded Model

After removing systemic confounders, an oral-specific signature persisted with moderate diagnostic accuracy, primarily driven by dental damage and stimulated salivary flow. To determine if a disease-specific oral signature could be identified, we trained a model restricted to oral phenotyping variables after exclusion of systemic confounders. Removal of these confounders resulted in a decrease in model performance, with the AUC dropping from 0.92 to 0.82 (Fig. 4A). Despite this reduction, the oral-only model outperformed the systemic-only baseline (AUC 0.78).

In the absence of systemic data, accumulated dental damage, captured by the DMFT index, emerged as the strongest individual predictor according to SHAP values, followed by stimulated salivary flow and the absence of dental prostheses (Fig. 4B, C). While the biological model showed robust global performance, a stratified analysis of accuracy by age revealed a non-linear trajectory. The model achieved peak precision of 83.3% in the 50–60 age range, corresponding to the typical onset of established disease. However, accuracy dropped to 66.7% in patients over 70 years (Fig. 4D). This age-dependent loss points to a clinical overlap in geriatric populations, prompting a further multivariable investigation to disentangle whether this severe dysfunction is driven by biological aging or the autoimmune process itself (Fig. 4E). This age-dependent loss of accuracy points to a clinical overlap in geriatric populations, prompting further investigation to disentangle whether this severe dysfunction is driven by biological aging or the autoimmune process itself.

3.5 Independent Associations with Stimulated Salivary Flow

A central question in Sjögren's research is whether hyposalivation is driven by biological aging, menopause, or the autoimmune process itself. To evaluate this, our analysis provides three converging lines of evidence focusing on: (a) the hormonal independence of the functional decline, (b) the existence of divergent age-related damage trajectories, and (c) a quantitative hierarchy of adjusted associations.

Multivariate regression analysis revealed distinct damage trajectories (Fig. 5C). Controls exhibited a steeper age-related decline (-0.025 mL/min/year) and a high theoretical intercept (2.50 mL/min). In contrast, the Sjögren group yielded a shallow regression slope (-0.013 mL/min/year) with a low intercept (1.45 mL/min).

When stratifying by menopausal status within a controlled age window of 45–55 years, Sjögren's patients exhibited severe hyposalivation even in the premenopausal stage ($n = 16$), with a mean flow of

0.44 mL/min versus 1.13 mL/min in premenopausal controls (n = 19) (Fig. 5A). The correlation between years since menopause and salivary flow across the cohort was weak (Pearson $r=-0.216$, $p = 0.016$; Spearman $\rho=-0.206$, $p = 0.022$) (Fig. 5B).

We quantified the independent contribution of each determinant using a multivariate linear regression model ($R^2 = 0.187$, $p < 0.001$). The diagnosis of Sjögren's syndrome was the strongest predictor, causing an immediate reduction of 0.52 mL/min (95% CI [-0.73, -0.29], $p < 0.001$). The use of hyposalivation-inducing medication caused a comparable reduction of 0.45 mL/min (95% CI [-0.75, -0.15], $p = 0.003$). By contrast, biological aging contributed a modest decline of 0.018 mL/min per year (95% CI [-0.027, -0.010], $p < 0.001$) (Fig. 5D, E). These findings suggested a hierarchy of damage where autoimmune pathology and medication drastically outweigh the contribution of biological aging to salivary dysfunction.

4. Discussion

Traditional strategies for addressing confounding in observational studies, such as propensity score matching or standard multivariate adjustments, often encounter limitations in high-dimensional clinical datasets. These methods may reduce sample size or overlook non-linear dependencies, potentially compromising generalizability[9–11]. Our cohort represents a real-world data scenario characterized by structured, imbalanced, and phenotypically entangled data. This high-dimensional setting appears to create a feature space distortion, where machine learning models, optimizing for predictive accuracy rather than causal validity, may preferentially exploit high-variance, easy-to-detect confounders rather than subtle biological signals. To address this, our approach sought to move beyond simple p -value-based adjustment by implementing a multi-stage computational framework. The main contribution of the study is not merely a classifier for Sjögren's Syndrome, but a computational framework for distinguishing disease-related signals from structured clinical noise.

4.1. Metabolic Confounding as a Driver

Our unsupervised analysis highlighted a significant phenotypic overlap between the cohorts, driven by a metabolic and pharmacological confounding profile in the control group rather than oral pathology alone. While standard machine learning models exploited this clustering pattern, characterized by diabetes, cardiovascular risk, and polymedication, as a diagnostic shortcut, this computational behavior reflects a profound etiological divergence. In literature, conditions such as diabetes are known to induce hyposalivation through autonomic neuropathy, glandular inflammatory infiltrates, and metabolic disturbances[37]. Similarly, polypharmacy frequently exerts well-documented xerogenic side effects, particularly in older adults[38]. Interestingly, our analysis revealed a negative correlation between smoking and the Sjögren's group. This "Tobacco Paradox" may stem from a behavioral adaptation: the additional mucosal irritation caused by smoke could be intolerable for patients with compromised glandular function, prompting earlier cessation[39]. Consequently, the label of "healthy control" in Sjögren's studies can be highly misleading. Our findings emphasize that metabolically compromised

controls present an alternative, iatrogenic pathway to xerostomia, which must be actively disentangled to prevent diagnostic algorithms from learning systemic comorbidities instead of the true autoimmune signature.

4.2. The Hierarchy of Oral Damage

By mathematically isolating the specific weight of each factor, our analysis established a clear hierarchy of damage where autoimmune pathology drastically outweighs both medication use and biological aging. Rather than a progressive decline, the presence of Sjögren's Syndrome behaves as a dominant floor effect, causing a massive, early-onset baseline loss of salivary function. Biologically, this is likely to reflect the irreversible destruction of acinar tissue by lymphocytic infiltration during the early stages of the disease[6], contrasting with the linear, gradual physiological atrophy associated with natural senescence.

Furthermore, our data revealed no significant difference in salivary flow between pre- and postmenopausal women, suggesting that the transition itself is not the trigger. Moreover, the weak correlation found between years since menopause and salivary flow appears to be driven by chronological age rather than shifting hormonal status. In contrast, classical literature frequently suggests that hormonal changes during the climacteric period contribute significantly to oral dryness and decreased salivary flow[40–42]. These studies often link the activity of the estrogen receptor in the salivary glands with the subjective and objective symptoms of xerostomia observed in aging women. Ultimately, these findings indicate that autoimmune-mediated glandular damage operates independently of reproductive aging, suggesting that severe hyposalivation is driven by a distinct, early autoimmune event rather than menopausal transition.

4.3. Validation of the Intrinsic Oral Signature

While our initial model achieved an elevated AUC (0.92), this performance was inflated by “shortcut learning” where the algorithm prioritized a metabolic phenotype over autoimmune damage. By removing these shortcuts, we identified a de-confounded oral signature of hyposalivation and dental damage that retained robust diagnostic power (AUC 0.82). This transition illustrates a feature masking effect, where the intrinsic biological signal only became prominent once the dominant informational gain of metabolic confounders was ablated. In the literature, this behavior is a recognized vulnerability in medical machine learning; classifiers often optimize for easy-to-detect confounding variables rather than complex disease recognition[35, 43]. High predictive accuracy can mask causal invalidity; however, by actively de-confounding the feature space, we confirmed that a unique, intrinsic oral signature persists independently of systemic comorbidities.

Finally, a notable clinical paradox emerged regarding rehabilitation: despite having significantly worse dental status, Sjögren's Syndrome patients demonstrated lower prosthesis usage compared to controls. This negative correlation is clinically coherent, as reduced salivary flow impairs the capillary adhesion required for denture retention while increasing friction and mucosal pain[44]. While literature suggests that implant-supported prostheses are a viable alternative to overcome these functional challenges[45],

the high systemic disease burden in this population often leads to under-referral[46]. This paradox highlights a significant unmet need in oral rehabilitation, where the very symptoms that necessitate treatment simultaneously act as a barrier to conventional prosthetic success.

4.4. The Diagnostic Convergence Zone in Geriatrics

Our age-stratified analysis identified a biological limit for oral diagnosis; over 70 years we found a lack of significant difference in salivary flow between cohorts and a drop in model precision. In the literature, this process is attributed to the synergistic effect of senile glandular atrophy, accumulation of dental damage, autonomic neuropathy from metabolic disorders, and cumulative anticholinergic burden of polypharmacy[5, 7]. This phenomenon represents a phenotypic convergence, in which age-related oral deterioration becomes indistinguishable from the damage associated with Sjögren's Syndrome, making oral clinical signs nonspecific in the geriatric population.

4.5. Limitations and Future Work

This study has several limitations that warrant consideration. First, the cross-sectional design precludes causal inference about trajectories of salivary decline; longitudinal data would be required to confirm whether the observed differences in intercepts and slopes represent true divergent pathways. Second, the Sjögren's syndrome and control cohorts were recruited from different clinical settings (rheumatology vs. gynecology clinics), introducing potential selection bias that may have contributed to the higher comorbidity burden in controls. Third, specific dosage data for xerogenic medications were not available, which may introduce residual confounding.

The absence of external validation limits generalizability. The metabolic structure identified in this cohort may not replicate in populations with different comorbidity patterns or healthcare systems. Multi-centric studies with harmonized recruitment protocols are needed to evaluate the broader epidemiological generalizability of these findings.

Finally, our oral model relies exclusively on clinical and oral features. The integration of proteomic or transcriptomic data could potentially further refine the intrinsic disease signature and resolve the diagnostic overlap observed in elderly patients. Despite these limitations, the methodological lessons regarding confounding structure and shortcut learning are likely applicable to other complex, multimorbid conditions where disease signals may be obscured by physiological aging, polypharmacy, and systemic comorbidities.

5. Conclusions

In this observational cohort, apparent diagnostic performance in Sjögren's syndrome was substantially influenced by confounding factors related to comorbidity and medication burden. An oral-specific signature persisted after excluding systemic variables but showed reduced accuracy in older patients due to phenotypic convergence.

For clinicians evaluating xerostomia, these findings underscore the importance of comprehensive medication review and comorbidity assessment before attributing symptoms to autoimmune disease or aging. For researchers, they highlight the need to audit prediction models for reliance on systemic confounders and to carefully characterize control groups. In geriatric populations where phenotypic overlap is common, oral clinical signs alone may lack specificity, supporting a multimodal diagnostic approach.

Declarations

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Use of Generative AI

Claude was also used to assist with formatting and documentation of the experimental code. All content was written by the authors, and the final text was carefully reviewed to ensure accuracy and clarity.

CRedit authorship contribution statement

Kristina Lacasta: Conceptualization, Technical Methodology, Technical analysis, Writing – Original Draft. Angela Gandara-Alvarez: Conceptualization, Clinical Methodology, Clinical analysis, Writing – Original Draft. Maria J. Rus: Patient Evaluation. Cristiane Cantiga-Silva: Patient Evaluation. Virginia Moreira Navarrete: Patient Recruitment. Carmen Dominguez Quesada: Patient Recruitment, Jose Javier Perez Venegas: Patient Evaluation. Juan Antonio Ortega: Supervision, Writing – Review & Editing. Aurea Simon-Soro: Conceptualization, Supervision, Writing – Review & Editing, Funding Acquisition.

Funding

This work was supported by the Spanish Ministry of Science and Innovation Grant ID PID2020-118557GA-I00 funded by MICIU/AEI/10.13039/501100011033 (A.S.-S.).

Author Contribution

K.L: Conceptualization, Technical Methodology, Technical analysis, Writing – Original Draft. A.G-A.: Conceptualization, Clinical Methodology, Clinical analysis, Writing – Original Draft. M J.R.: Patient Evaluation. C.C-S.: Patient Evaluation. V.M.N.: Patient Recruitment. C.D.Q.: Patient Recruitment.JJ.P.V.:

Acknowledgement

This work was supported by the Spanish Ministry of Science and Innovation Grant ID PID2020-118557GA-I00. The authors would like to express their deepest gratitude to all the patients who participated in this study. We also thank the clinical and administrative staff at the Virgen Macarena University Hospital (Seville, Spain) for their assistance in patient recruitment and data collection.

Data Availability

All code and analysis are publicly available on GitHub:

https://github.com/krilaclop/HormoBiome_Confounding_Factors_Sjogren.git. De-identified data summaries and model configurations are provided within the repository to support the reproducibility of the findings. Raw clinical data are restricted due to institutional patient privacy regulations and ethical requirements.

References

1. Brito-Zerón P, Baldini C, Bootsma H et al (2016) Sjögren syndrome. *Nat Rev Dis Primers* 2:16047. <https://doi.org/10.1038/nrdp.2016.47>
2. Bruno KA, Morales-Lara AC, Bittencourt EB et al (2022) Sex differences in comorbidities associated with Sjögren's disease. *Front Med (Lausanne)* 9:958670. <https://doi.org/10.3389/fmed.2022.958670>
3. McCoy SS, Sampene E, Baer AN (2020) Association of Sjögren's Syndrome With Reduced Lifetime Sex Hormone Exposure: A Case-Control Study. *Arthritis Care Res (Hoboken)* 72:1315–1322. <https://doi.org/10.1002/acr.24014>
4. Pijpe J, Kalk WWI, Bootsma H et al (2007) Progression of salivary gland dysfunction in patients with Sjögren's syndrome. *Ann Rheum Dis* 66:107–112. <https://doi.org/10.1136/ard.2006.052647>
5. Coelho CPES, Da Silva JR, Marques RCR et al (2026) Salivary flow and aging: a cross-sectional estimate of the annual decline per year of age. *Oral Surgery, Oral Medicine. Oral Pathol Oral Radiol* 141:44–53. <https://doi.org/10.1016/j.oooo.2025.08.008>
6. Shiboski CH (2017) others 2016 ACR/EULAR classification criteria for primary Sjögren's syndrome. *Arthritis & Rheumatology* 69:35–45. <https://doi.org/10.1002/art.39859>
7. Boonyoung K, Buajeeb W, Panpradit N et al (2026) Prevalence of xerostomia and hyposalivation in older adults: a systematic review and meta-analysis. *BMC Oral Health* 26:1125. <https://doi.org/10.1186/s12903-026-08579-x>

8. Da Silva JR, Melo CDCDO, Leite EC et al (2026) Machine-learning analysis of short-term within-person salivary flow variation in a longitudinal cohort. *Arch Oral Biol* 191:106731. <https://doi.org/10.1016/j.archoralbio.2026.106731>
9. Padmanabhan K, Lu M, Feng D et al (2026) Explainable machine learning in healthcare: methods, interpretation, and applications for clinical research. *J Am Med Inform Assoc* 33:1510–1522. <https://doi.org/10.1093/jamia/ocag077>
10. Shiddik MAB (2026) Explainable Artificial Intelligence in Healthcare: Current Landscape, Challenges, and Future Directions. *Health Sci Rep* 9:e72172. <https://doi.org/10.1002/hsr2.72172>
11. Zhang K, Wang D, Lin F et al (2026) A comprehensive review of explainable artificial intelligence in healthcare methods, evaluation, and clinical integration. *iScience* 29:115026. <https://doi.org/10.1016/j.isci.2026.115026>
12. Alvariño C, others Bagan L (2021) Stimulated whole salivary flow rate in Sjögren syndrome. *Medicina Oral, Patología Oral y Cirugía Bucal* 26:e404–e407. <https://doi.org/10.4317/medoral.24736>
13. Da Silva JR, Marques RCR, Nunes FPS et al (2026) What Is the Normal Salivary Flow Rate in Healthy Adults? A Systematic Review with Meta-Analyses. *JDR Clin Translational Res* 11:117–132. <https://doi.org/10.1177/23800844251336205>
14. Dhingra K, Vandana KL (2011) Indices for measuring periodontitis: a literature review. *Int Dent J* 61:76–84. <https://doi.org/10.1111/j.1875-595X.2011.00018.x>
15. World Health Organization (2013) Oral health surveys: Basic methods, 5th edn. World Health Organization, Geneva, Switzerland
16. Shulman JD, Cappelli DP (2008) Epidemiology of dental caries. In: Cappelli DP, Mobley CC (eds) *Prevention in Clinical Oral Health Care*. Mosby Elsevier, St. Louis, MO
17. The pandas development team (2020) pandas-dev/pandas: Pandas
18. Harris CR, Millman KJ (2020) others Array programming with NumPy. *Nature* 585:357–362. <https://doi.org/10.1038/s41586-020-2649-2>
19. Virtanen P, Gommers R (2020) others SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* 17:261–272. <https://doi.org/10.1038/s41592-019-0686-2>
20. Pedregosa F, Varoquaux G (2011) others Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830
21. Hunter JD (2007) Matplotlib: A 2D graphics environment. *Comput Sci Eng* 9:90–95. <https://doi.org/10.1109/MCSE.2007.55>
22. Waskom ML (2021) seaborn: statistical data visualization. *J Open Source Softw* 6:3021. <https://doi.org/10.21105/joss.03021>
23. Van Buuren S, Groothuis-Oudshoorn K (2011) mice: Multivariate imputation by chained equations in R. *J Stat Softw* 45:1–67. <https://doi.org/10.18637/jss.v045.i03>

24. White IR, Royston P, Wood AM (2011) Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 30:377–399. <https://doi.org/10.1002/sim.4067>
25. Greenacre M, Groenen PJF (2022) others Principal component analysis. *Nature Reviews Methods Primers* 2:100. <https://doi.org/10.1038/s43586-022-00184-w>
26. Gewers FL (2021) others Principal component analysis: A natural approach to data exploration. *ACM Computing Surveys* 54:. <https://doi.org/10.1145/3447755>
27. Trujillano J, Serviá L (2025) others Methodological review of classification trees for risk stratification. *Nutrients* 17:1903. <https://doi.org/10.3390/nu17111903>
28. Quist J, Taylor L (2021) others Random forest modelling of high-dimensional mixed-type data for breast cancer classification. *Cancers* 13:991. <https://doi.org/10.3390/cancers13050991>
29. Hameed I, Sharpe S, Barcklow D (2022) others BASED-XAI: Breaking ablation studies down for explainable artificial intelligence
30. Merrick L, Taly A (2019) Randomized ablation feature importance
31. Nohara Y, Matsumoto K, Soejima H, Nakashima N (2022) Explanation of machine learning models using Shapley Additive Explanation and application for real data in hospital. *ICT Express* 8:184–188. <https://doi.org/10.1016/j.icte.2021.05.001>
32. Lundberg S (2021) SHAP (Shapley additive explanations) documentation
33. Seabold S, Perktold J (2010) statsmodels: Econometric and statistical modeling with Python. In: 9th Python in Science Conference
34. Fumo-Dos-Santos C, Ferreira JC (2023) Dealing with confounding in observational studies. *Jornal Brasileiro de Pneumologia* 49:e20230281. <https://doi.org/10.36416/1806-3756/e20230281>
35. Assimon MM (2021) Confounding in observational studies evaluating the safety and effectiveness of medical treatments. *Kidney360* 2:1156–1159. <https://doi.org/10.34067/KID.0007022020>
36. Wallach JD, Serghiou S, Chu L (2020) others Evaluation of confounding in epidemiologic studies assessing alcohol consumption on the risk of ischemic heart disease. *BMC Medical Research Methodology* 20:64. <https://doi.org/10.1186/s12874-020-0914-6>
37. Mata AD, Marques D (2004) others Effects of diabetes mellitus on salivary secretion and its composition in the human. *Molecular and Cellular Biochemistry* 261:137–142. <https://doi.org/10.1023/B:MCBI.0000028748.40917.6f>
38. Barbe AG (2018) Medication-induced xerostomia and hyposalivation in the elderly. *Drugs Aging* 35:877–885. <https://doi.org/10.1007/s40266-018-0588-5>
39. Dietrich T, Ower P, Tank M et al (2019) Periodontal diagnosis in the context of the 2017 classification system of periodontal diseases and conditions – implementation in clinical practice. *Br Dent J* 226:16–22. <https://doi.org/10.1038/sj.bdj.2019.3>
40. Meurman JH, Tarkkila L, Tiitinen A (2009) The menopause and oral health. *Maturitas* 63:56–62. <https://doi.org/10.1016/j.maturitas.2009.02.009>

41. Minicucci E, Pires R, Vieira R et al (2013) Assessing the impact of menopause on salivary flow and xerostomia. *Aust Dent J* 58:230–234. <https://doi.org/10.1111/adj.12057>

42. Frutos R, Rodríguez S, Miralles-Jorda L, Machuca G (2001) Oral manifestations and dental treatment in menopause. *Med Oral* 7:26–35

43. Ferrari E, Retico A, Bacciu D (2020) Measuring the effects of confounders in medical supervised classification problems: the confounding index (CI). *Artif Intell Med* 103:101804. <https://doi.org/10.1016/j.artmed.2020.101804>

44. Albrecht K, Callhoff J, Westhoff G et al (2016) The prevalence of dental implants and related factors in patients with Sjögren syndrome: Results from a cohort study. *J Rheumatol* 43:1380–1385. <https://doi.org/10.3899/jrheum.151167>

45. Almeida D, Vianna K (2017) others Dental implants in Sjögren’s syndrome patients: A systematic review. *PLOS ONE* 12:e0189507. <https://doi.org/10.1371/journal.pone.0189507>

46. Fox RI, Fox CM (2021) others Treatment of Sjögren’s syndrome: current therapy and future directions. *Rheumatology* 60:2066–2074. <https://doi.org/10.1093/rheumatology/kez142>

Figures

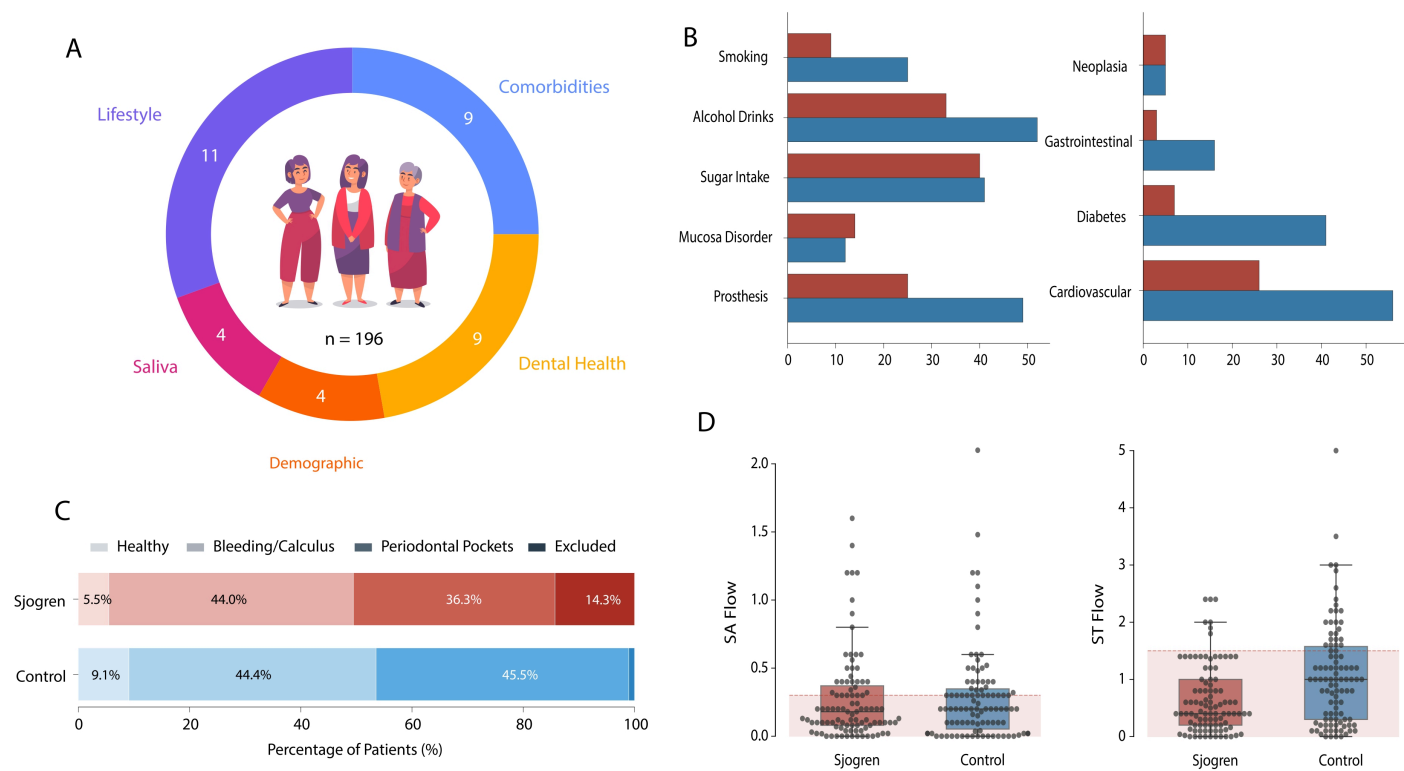


Figure 1

Clinical and demographic characterization of the study cohort. (A) Distribution of the 37 collected variables grouped into five thematic categories: Comorbidities (9 variables), Dental Health (9 variables),

Demographics (4 variables), Saliva (4 variables), and Lifestyle (11 variables). **(B)** Comparison of lifestyle, oral health, and comorbidity variables between Sjögren's syndrome patients (red) and controls (blue). Cardiovascular disease was the most prevalent comorbidity in both groups. **(C)** Distribution of periodontal health status by Community Periodontal Index, expressed as percentage of patients. Categories: Healthy, Bleeding/Calculus, Periodontal Pockets, and Excluded (Sjögren's group only). **(D)** Salivary flow rates between Sjögren's patients (red) and controls (blue), with individual data points overlaid. Left: unstimulated salivary flow (SA_Flow, mL/min). Right: stimulated salivary flow (ST_Flow, mL/min). Red dashed horizontal lines indicate clinical hyposalivation thresholds (SA ≤ 0.1 mL/min; ST ≤ 1.5 mL/min).

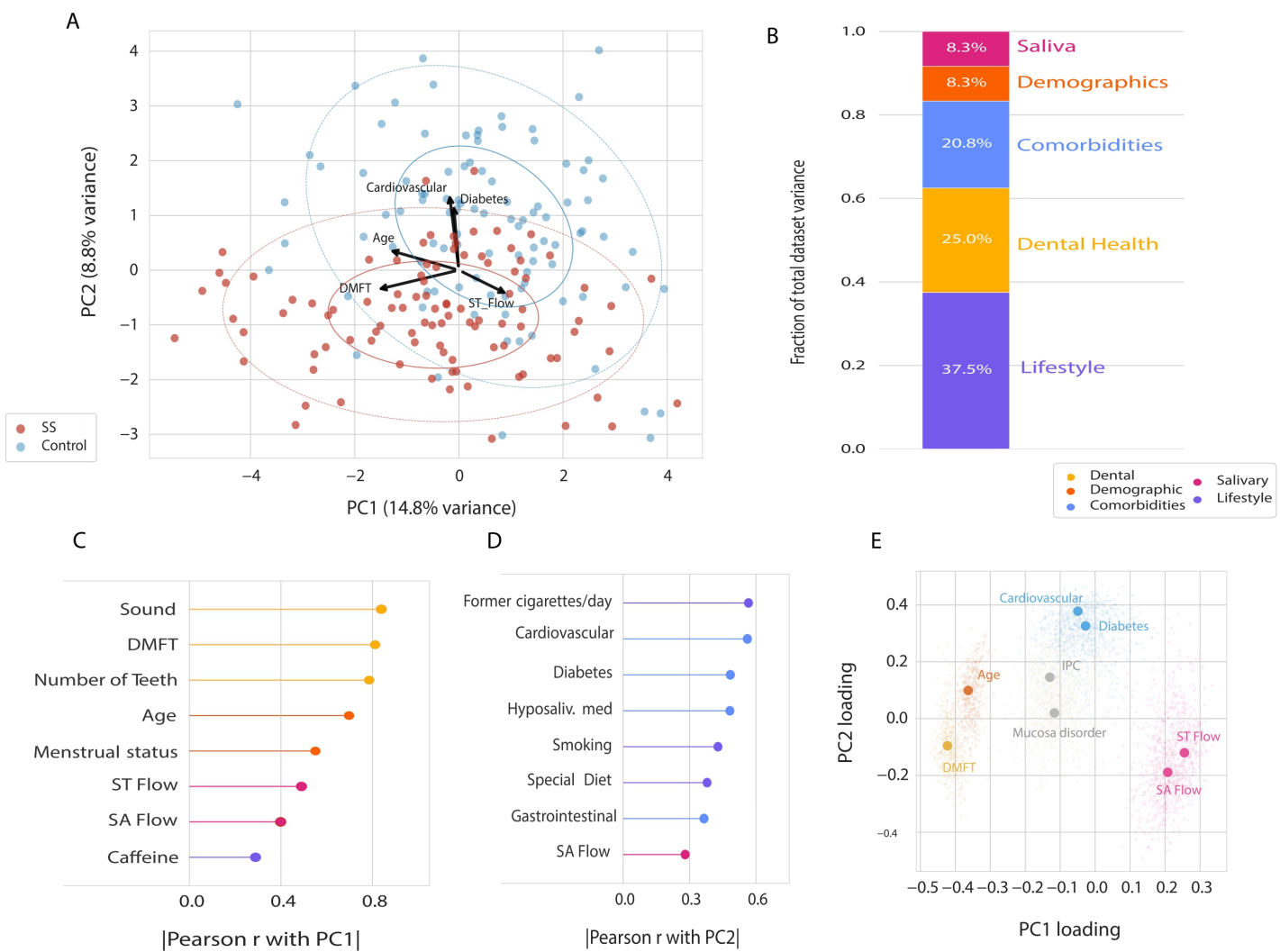


Figure 2

Principal component analysis of clinical and oral variables. **(A)** PCA biplot showing PC1 (14.8% of variance) and PC2 (8.8%). Sjögren's patients (red) and controls (blue) overlap substantially, indicating absence of clean diagnostic separation. Loading arrows show direction and magnitude of representative variables: Age and DMFT load primarily on PC1, while cardiovascular disease and Diabetes load on PC2.

(B) PC1 loading profile (oral-aging axis). Pearson correlations between each variable and PC1 scores, ranked by absolute magnitude. PC1 is dominated by dental deterioration proxies (Sound teeth $r=0.84$, DMFT $r=0.81$, Number of Teeth $r=0.79$) and Age ($r=0.70$). **(C)** PC2 loading profile (metabolic-pharmacological axis). PC2 is driven by lifestyle and comorbidity variables enriched in controls: former cigarette consumption ($r=0.57$), cardiovascular disease ($r=0.56$), diabetes ($r=0.48$), hyposalivation medication ($r=0.48$), and active smoking ($r=0.43$). **(D)** Bootstrap loading stability. PCA refit on 500 bootstrap resamples. Each point represents PC1–PC2 loading pair from one resample; cloud spread reflects sampling uncertainty. Stability assessed as signal-to-noise ratio ($\text{SNR} = \text{dominant loading} / \text{SD}$; threshold $\text{SNR} \geq 2.5$). Age and DMFT show tightest clouds ($\text{SNR} > 18$). Cardiovascular disease and Diabetes stable on PC2 ($\text{SNR} > 3$). **(E)** Variance decomposition by variable group. Fraction of total dataset variance attributable to each variable group computed using loading²-weighted attribution across all PCA components. Lifestyle variables account for largest share (37.5%), followed by Dental Health (25.0%) and Comorbidities & Medication (20.8%).

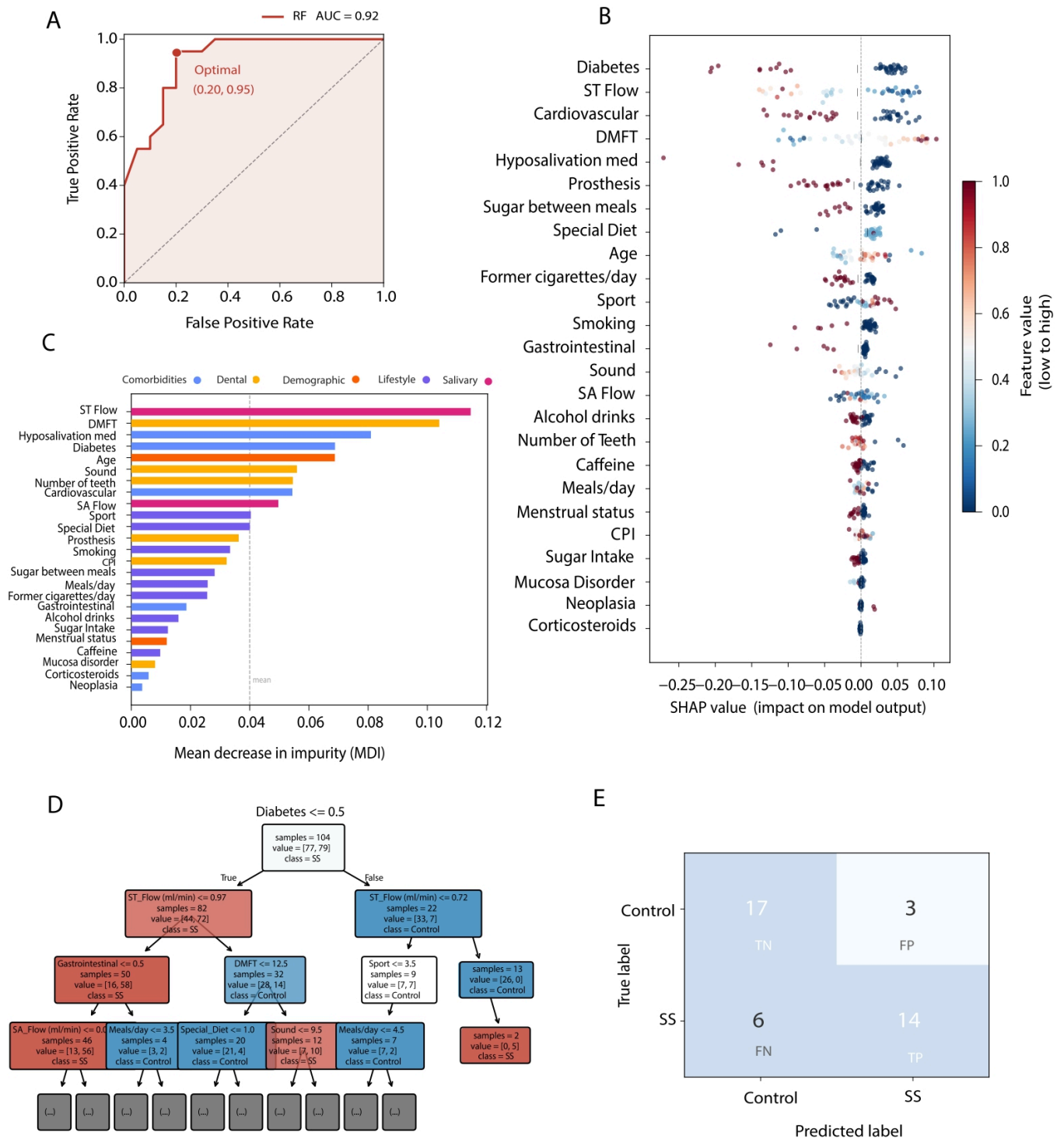


Figure 3

Random Forest classifier performance and interpretability on full dataset. (A) ROC curve with optimal threshold selected by Youden's J statistic. **(B)** SHAP beeswarm plot: each dot represents one test sample; colour indicates feature value (red=high, blue=low). Salivary flow (ST_Flow, SA_Flow) is dominant discriminator. **(C)** Mean Decrease in Impurity (MDI) feature importances confirming ranking. **(D)** First tree of ensemble, pruned to depth 3. Diabetes emerges as root node, acting as primary

exclusion criterion. **(E)**Confusion matrix on held-out test set with key diagnostic metrics (accuracy, sensitivity, specificity, positive and negative predictive values).

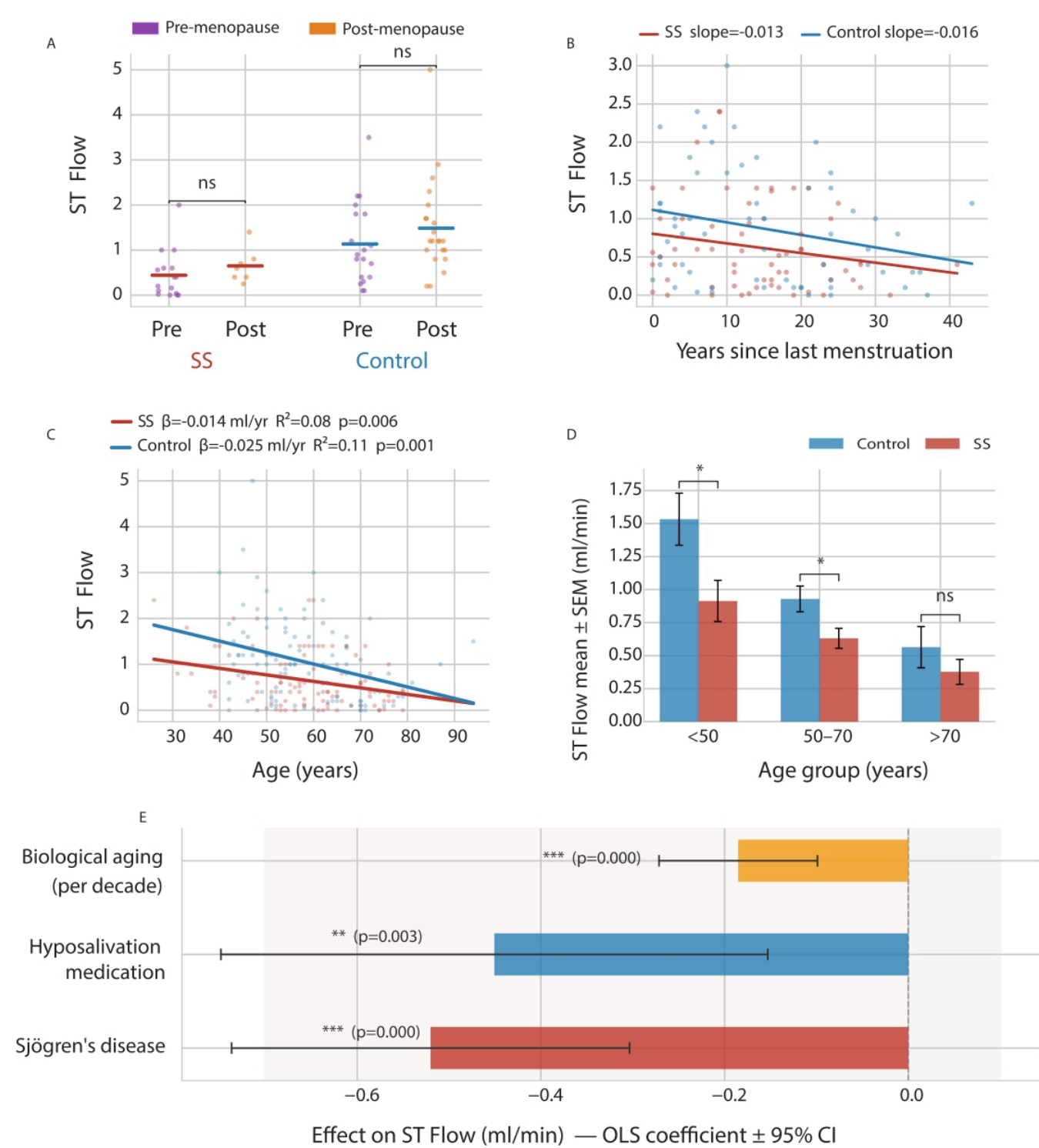


Figure 4

Oral-only model performance after confounder exclusion. (A) ROC comparison: full model (all features, dashed blue) vs oral-only model (confounders removed, solid red). **(B)**SHAP beeswarm for oral-only

model: each dot represents one test sample; colour indicates feature value (red=high). **(C)** Top 10 MDI importances for oral-only model; blue bars show feature importance of full model for comparison. **(D)** Oral-only model accuracy by age group (5-fold CV); dashed line indicates global accuracy. Peak precision in 50–60 age range; reduced accuracy in patients over 70 years. **(E)** Confusion matrix on held-out test set with key diagnostic metrics.

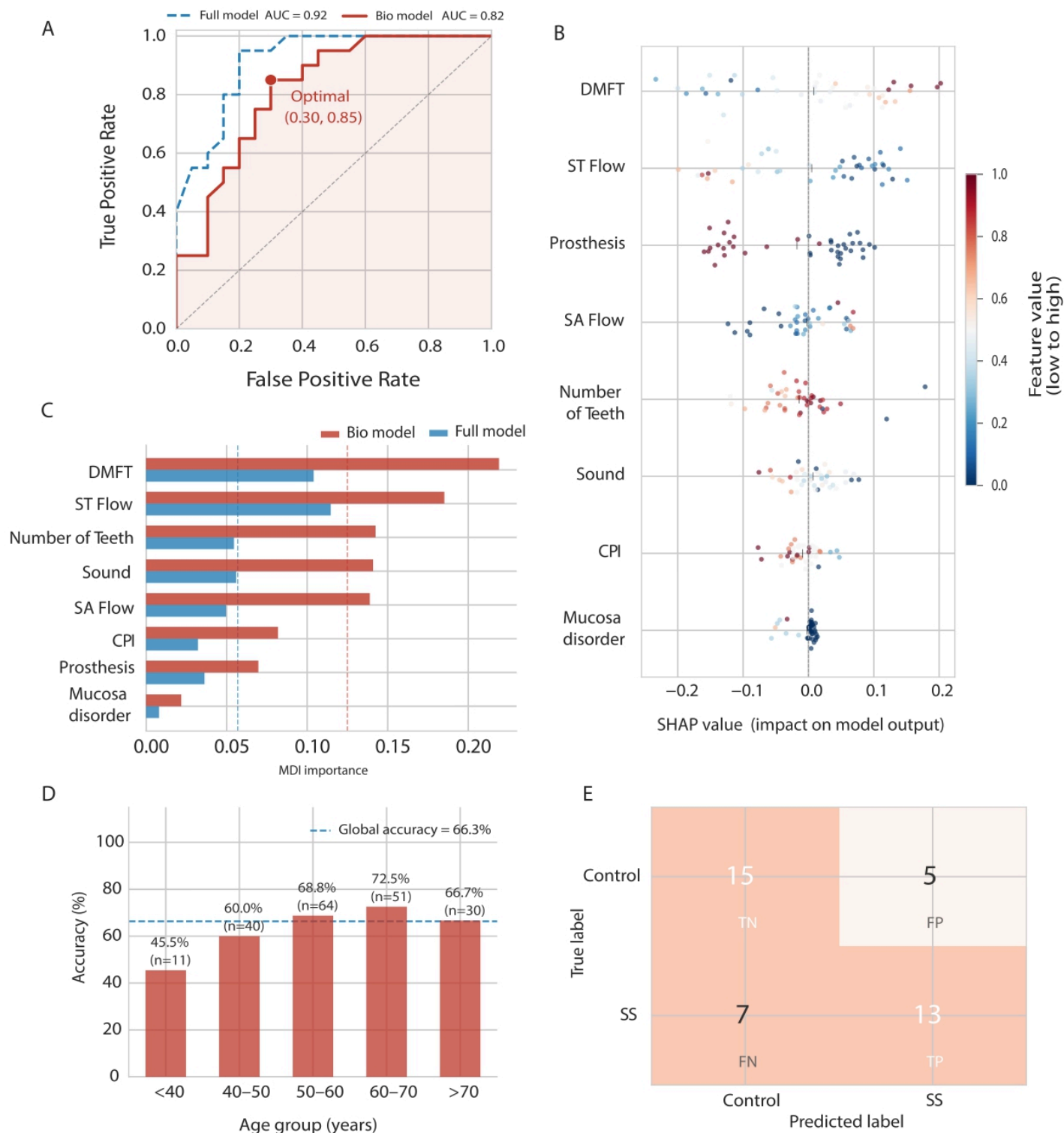


Figure 5

Determinants of stimulated salivary flow. (A) ST Flow in menopausal transition window (age 45–55): pre- vs post-menopause, stratified by diagnosis. Horizontal bars indicate group mean; dots show individual values. **(B)** Linear association between years since last menstruation and ST Flow (Pearson $r=-0.216$, $p=0.016$). **(C)** Age-dependent trajectories: regression slopes quantify deterioration rate per year. Controls show steeper decline (-0.025 mL/min/year) than Sjögren's patients (-0.013 mL/min/year). **(D)** Mean ST Flow $\pm$ SEM by age group. Significance: $p<0.001$, $p<0.01$, $*p<0.05$, ns=not significant. **(E)** Multivariate OLS regression decomposition (*ceteris paribus*). Sjögren's diagnosis associated with largest reduction in ST Flow, followed by xerogenic medication use. Biological aging shows modest annual decline.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTableS1.pdf](#)
- [Graphicalabstractv2.jpg](#)