Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Development and Internal Validation of a Self-Reported Oral Health Prediction Model for Periodontitis Risk Assessment

Alias Abd Aziz
University of Malaya
Tayf Naif R Alshammari
University of Malaya
Ainol Haniza Kherul Anuwar
University of Malaya
Nor Adinar Baharuddin
noradinar@um.edu.my
University of Malaya

Research Article

Keywords: internal validation, prediction model, periodontitis, risk assessment, self-reported oral health

Posted Date: November 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11168417/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

Background

Self-reported oral health measures may provide a practical approach for identifying individuals at risk of periodontitis where comprehensive periodontal examination is not routinely available. This study aimed to develop and internally validate a self-reported oral health (SROH)-based prediction model for periodontitis risk assessment.

Methods

The study included 191 adults who completed a six-item SROH questionnaire and had full-mouth periodontal examination records. Periodontitis was defined based on the 2018 AAP/EFP classification. An SROH-only model and a combined model incorporating demographic and behavioural predictors were developed using multivariable logistic regression and internally validated using 1,000 bootstrap resamples. Model performance was assessed using discrimination, calibration, classification and decision curve analysis.

Results

Of the 191 participants, 92 (48.2%) had periodontitis. Self-reported gum disease, loose teeth and alveolar bone loss were independently associated with periodontitis, while age was the only additional predictor retained in the combined model. Apparent AUROC increased from 0.714 for the SROH-only model to 0.763, with optimism-corrected AUROC values of 0.702 and 0.745, respectively. The combined model achieved a sensitivity of 70.7%, specificity of 74.7%, positive predictive value of 72.2% and negative predictive value of 73.3%. The combined model showed generally satisfactory calibration and greater net benefit across a range of threshold probabilities.

Conclusions

A prediction model incorporating self-reported gum disease, loose teeth, bone loss, and age demonstrated acceptable internally validated performance for identifying individuals at risk of periodontitis. The model may support non-invasive risk assessment and referral for periodontal examination in community and non-dental settings. External validation is required before broader implementation.

Background

Periodontitis is one of the most common chronic oral diseases worldwide. It causes progressive destruction of the tooth-supporting tissues and is a major cause of tooth loss among adults. Globally,

approximately 30–50% of adults are affected by periodontal disease, while 9–11% have severe forms of periodontal disease [1–3]. In Malaysia, the National Oral Health Survey of Adults (NOHSA), reported that 94% of dentate adults had some form of periodontal condition, while almost half had periodontal pockets. The mean number of healthy sextants was also low, at 0.68 per person, indicating a substantial need for periodontal assessment and care [4].

Although a full-mouth periodontal examination remains necessary for definitive diagnosis, it may not be feasible for population-level screening. Self-reported oral health (SROH) tools provide a practical and non-invasive approach to identifying individuals who may be at risk of having periodontitis and should be referred for further periodontal assessment. These questionnaires are inexpensive, easily administered and scalable which make them potentially suitable for community screening programs, primary health care settings, and other non-dental environments where full mouth periodontal examination cannot be routinely performed [5–8].

Previous studies have shown that combining several self-reported measures of periodontal symptoms, history and awareness performs better than using individual question alone to identify periodontitis [9]. Several prediction models have also incorporated readily available demographic and behavioural variables (e.g., age, smoking status) to improve the identification of individuals at risk of having periodontitis [6, 8, 10–11]. Such models offer an opportunity to incorporate periodontal risk assessment into community programmes and routine medical care, including primary care consultations and chronic disease clinics, to facilitate the referral of individuals who may require comprehensive periodontal assessment [7–8].

Despite the promising performance of self-reported periodontal screening models, their transferability across populations and healthcare settings remains a challenge. Differences in disease prevalence, demographic characteristics, oral health awareness, socioeconomic background, cultural perceptions, and access to dental care may influence how individuals understand and respond to self-reported questions [8–9]. For example, Nijland et al. [7] externally validated a previously developed screening model in a medical outpatient setting and reported lower discrimination for overall periodontitis than in the original study, although its ability to identify severe disease remained acceptable. These findings highlight the importance of developing and validating prediction models in populations that reflect their intended setting and users.

Reliable model development also requires an assessment of whether the apparent performance observed in the development sample may have been influenced by overfitting. Some previous studies have relied on apparent performance or split-sample validation which may produce unstable or overly optimistic estimates, particularly when the available sample is limited. Bootstrap resampling is recommended for internal validation because it allows the full sample to be used for model development while estimating and correcting for optimism in model performance [12–14]. External validation remains necessary to determine whether the model performs adequately in other populations and healthcare settings [7].

Evidence on SROH-based periodontitis prediction models remains limited in Malaysia. A recent Malaysian study demonstrated the potential of combining SROH measures with salivary biomarkers and machine-learning algorithms for non-invasive periodontal screening [15]. However, such an approach requires saliva collection, biomarkers analysis and additional technical resources that may limit its application in larger-scale community settings. A simpler prediction model based entirely on information that can be obtained without clinical examination or laboratory testing may be more feasible for community-level risk assessment. Furthermore, few studies have adopted a comprehensive model-development approach that includes internal validation, assessment of discrimination and calibration, evaluation of clinical usefulness, and translation of the final model into a clinically applicable tool [7, 16–17]. Therefore, this study aimed to develop and internally validate an SROH-based prediction model for identifying Malaysian adults at risk of having periodontitis and to translate the final model into a clinically applicable nomogram for non-invasive periodontal risk assessment.

Methods

Study Design and Participants

The prediction-model development study was reported in accordance with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement for model development with internal validation (Type 1b) [8]. The study linked self-reported oral health data collected specifically for the present study with existing full-mouth periodontal examination records from a parent study.

The study utilised data obtained from a previously conducted parent study involving 300 adults who were prospectively recruited between February and August 2024. Participants underwent full-mouth periodontal examination, sociodemographic data collection and saliva sampling. The parent study was approved by the Faculty of Dentistry Medical Ethics Committee, Universiti Malaya (Reference No.: DF RD2401/0010 (L)).

Between May and June 2025, participants from the parent study were invited to participate in the present study and completed a structured questionnaire that included the Malaysian SROH questionnaire. The questionnaire responses were linked to the participants' periodontal examination records using unique participant identifiers. The present study was approved under a separate research protocol by the Faculty of Dentistry Medical Ethics Committee, Universiti Malaya (Reference No.: DF OC2503/0073 (P)).

Participants were included if they: (i) were age ≥ 18 years; (ii) provided written informed consent for the present study; (iii) completed the structured self-reported periodontal screening questionnaire; and (iv) had complete full-mouth periodontal examination records from the parent study. Participants with incomplete questionnaire responses or incomplete clinical examination records were excluded. Of the

300 participants in the parent study, 191 (63.7%) agreed to participate and fulfilled the eligibility criteria for the present analysis.

Sample Size

All eligible participants from the parent study who completed the SROH questionnaire and had complete periodontal examination records were included. The sample size was therefore determined by the number of available participants who fulfilled the eligibility criteria rather than by an a priori sample size calculation. The final analysis included 191 participants, of whom 92 had periodontitis and 99 did not have periodontitis. Given the available sample size and number of outcome events, model complexity was considered during predictor selection to reduce the risk of overfitting. Bootstrap resampling was subsequently used to estimate and correct for optimism in model performance.[19]

Self-Reported Oral Health Variables

Six items from the Malaysian SROH questionnaire were evaluated as candidate predictors. The items captured participants' awareness, symptoms and previous experiences related to periodontal health. Participants were asked whether they had: (i) ever been told that they had gum disease; (ii) perceived the health of their gums as unhealthy; (iii) previously received periodontal treatment; (iv) ever experienced loose teeth; (v) ever been told by a dental professional that they had alveolar bone loss; or (vi) during the previous three months, noticed a tooth that did not look right, such as a shaky, tilted or drifted tooth.

The responses options for all six items were "yes," "no" and "don't know." For analysis, responses were recoded as binary variable, with "yes" classified as positive response, and "no," or "don't know," classified as a negative response. The questionnaire was administered in both English and Malay language using a self-administered format. The development and validation of the Malaysian SROH questionnaire have been undertaken separately, and the findings have been submitted for publication.

Demographic and Behavioural Variables

Demographic and behavioural characteristics considered as candidate predictors were age, sex, ethnicity, marital status, education level, monthly income, dental attendance pattern, presence of systemic disease, and smoking status. These variables were selected based on their established or potential associations with periodontitis. The information was obtained from the structured questionnaire administered in the present study. The categories and coding used for analysis are presented in the relevant tables.

Reference Standard and Periodontal Assessment

The reference standard was the full-mouth periodontal examination conducted during the parent study between February and August 2024. Trained and calibrated examiners used a standardised periodontal probe to measure probing pocket depth (PPD) and clinical attachment loss (CAL) at six sites per tooth.

Periodontitis was defined according to the 2018 Classification of Periodontal and Peri-Implant Diseases and Conditions developed by the American Academy of Periodontology (AAP) and the European Federation of Periodontology (EFP) [20]. A participant was classified as having periodontitis when interdental CAL was detectable at ≥ 2 non-adjacent teeth, or when buccal or oral CAL of ≥ 3 mm with PPD > 3 mm was detectable at ≥ 2 teeth, provided that the attachment loss could not attributed to non-periodontitis-related causes. Participants meeting the criteria for Stage I to Stage IV periodontitis were classified as having periodontitis. Periodontitis status was subsequently analysed as a binary outcome of periodontitis and non-periodontitis.

Examiner Calibration

Clinical examinations were performed by trained and calibrated examiners under the supervision of experienced periodontists. Prior to data collection, calibration exercises were conducted to standardise periodontal measurements. Inter-examiner and intra-examiner reliability for PPD and CAL were assessed using the Intraclass Correlation Coefficient (ICC), with values were > 0.75 for all parameters studied, indicating acceptable reliability.

Handling of Missing Data

Only participants with complete data for the candidate predictors and periodontitis outcome were included in the analysis. Participants with incomplete questionnaire responses or periodontal examination were excluded. No data imputation was performed.

Statistical Analysis and Model Development

Statistical analyses were conducted using the IBM SPSS Statistics version 29 (IBM Corp., Armonk, NY, USA) and R statistical software (R Foundation for Statistical Computing, Vienna, Austria). Statistical significance was set at $p < 0.05$.

Bivariable Analysis

Participant characteristics and SROH responses were compared according to periodontitis status using Pearson's chi-square test. The strength of association was assessed Cramér's V, while crude odds ratios (cORs) and corresponding 95% confidence intervals (CIs) were estimated using binary logistic regression. Variables with a $p < 0.25$ in the bivariable analysis were considered for inclusion in the multivariable model.

Multivariable Logistic Regression Modelling

Two multivariable logistic regression models were developed. Model 1 included only SROH variables identified as potential predictors in the bivariable analysis. Model 2 combined the SROH variables with eligible demographic and behavioural predictors. Adjusted odds ratios (aORs), corresponding 95% CIs and regression coefficients were reported for predictors retained in the final models. Multicollinearity among candidate predictors was assessed before finalising the models.

Internal Validation

The prediction models were internally validated using 1,000 bootstrap resamples with the rms package in R [21–23]. In each resample, the model was fitted and its performance was evaluated in both the bootstrap sample and original dataset. The mean difference in performance represented the estimated optimism, which was subtracted from the apparent performance to obtain optimism-corrected estimates. This approach allowed the full sample to be used for model development while quantifying potential overfitting [12–13].

Model Performance Assessment

Discrimination

Model discrimination was assessed by the area under the receiver operating characteristic curve (AUROC). The apparent and optimism-corrected AUROC values were reported for both prediction models, with corresponding 95% confidence intervals where available. Higher AUROC values indicated a greater ability to distinguish between participants with and without periodontitis.

Calibration

Model calibration was assessed using bootstrap-corrected calibration plots based on 1,000 bootstrap resamples. Predicted probabilities were plotted against the observed proportion of periodontitis and compared with the ideal 45° line. The plots were examined for systematic overestimation or underestimation of the probability of periodontitis.

Classification Performance

Predicted probabilities were converted into binary classifications using the optimal probability thresholds determined by the Youden Index. At the selected threshold, model performance was assessed using sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Bootstrap-derived 95% CIs were estimated for the AUROC, sensitivity and specificity.

Clinical Utility

Decision Curve Analysis (DCA) was conducted to evaluate the potential clinical usefulness of the prediction models. Net benefit was assessed across a range of threshold probabilities and compared with the default strategies of referring all participants and referring none. A model was considered

potentially useful when it provided a greater net benefit than both default strategies within a clinically relevant threshold range.

Model Presentation and Clinical Translation

The final combined model was presented as a logistic regression equation and translated into a nomogram using the *rms* package in R. Regression coefficients were converted into a point-based scoring system to facilitate estimation of an individual's probability of having periodontitis based on their predictor profile. The nomogram was developed to support interpretation and potential clinical application of the model, subject to further external validation.

Results

A total of 191 participants were included, of whom 92 (48.2%) were classified as having periodontitis and 99 (51.8%) as not having periodontitis. Participants with periodontitis were generally older and had lower educational attainment than those without periodontitis. Periodontitis status was also associated with marital status, while dental attendance showed a borderline association. No significant associations were observed for sex, ethnicity, monthly income, systemic disease or smoking status (Table 1).

Table 1
Demographic and Baseline Clinical Characteristics

Variables	All Participants	Non-Periodontitis n (%)	Periodontitis n (%)	p-value	Cramer's V
Total, n (%)	191 (100.0)	99 (51.8)	92 (48.2)		
Age category					
Below 40	145 (75.9)	87 (60.0)	58 (40.0)	**0.001**	0.290
40 And above	46 (24.1)	12 (26.1)	34 (73.9)		
Gender					
Male	48 (25.1)	24 (50.0)	24 (50.0)	0.769	0.021
Female	143 (74.9)	75 (52.4)	68 (47.6)		
Ethnicity					
Malay	134 (70.2)	65 (48.5)	69 (51.5)	0.227	0.125
Chinese	39 (20.4)	25 (64.1)	14 (35.9)		
Others	18 (9.4)	9 (50.0)	9 (50.0)		
Marital status					
Divorced	7 (3.7)	2 (28.6)	5 (71.4)	**0.001**	0.288
Married	81 (42.4)	30 (37.0)	51 (63.0)		
Single	103 (53.9)	67 (65.0)	36 (35.0)		
Education level					
Below Tertiary	18 (9.4)	4 (22.2)	14 (77.8)	**0.008**	0.191
Tertiary	173 (90.6)	95 (54.9)	78 (45.1)		
Monthly income					
< RM5,250	108 (56.5)	54 (50.0)	54 (50.0)	0.563	0.042
> RM 11,819	25 (13.1)	12 (48.0)	13 (52.0)		
RM 5250–RM11,819	58 (30.4)	33 (56.9)	25 (43.1)		
Dental Check-up					
Yes	68 (35.6)	41 (60.3)	27 (39.7)	0.082	0.126
No	123 (64.4)	58 (47.2)	65 (52.8)		

Variables	All Participants	Non-Periodontitis n (%)	Periodontitis n (%)	p-value	Cramer's V
Systemic disease					
Yes	20 (10.5)	10 (50.0)	10 (50.0)	0.862	0.013
none	171 (89.5)	89 (52.0)	82 (48.0)		
Ever Smoking					
Never	185 (96.9)	97 (52.4)	88 (47.6)	0.357	0.067
Yes	6 (3.1)	2 (33.3)	4 (66.7)		

Chi-square test: Bold indicates significance at p-value < 0.05.

Of the six SROH variables evaluated, four were significantly associated with clinically diagnosed periodontitis (Table 2). Self-reported alveolar bone loss showed the strongest association. Participants who reported having been told by a dental professional that they had alveolar bone loss had 8.89 times the odds of having periodontitis compared with those who did not report alveolar bone loss (cOR = 8.89; 95% CI: 2.54–31.07; p = 0.001).

Table 2
Univariate Associations of Self-Reported Oral Health Items

SROH Predictor Variable	Overall n (%)	Non-Periodontitis n (%)	Periodontitis n (%)	Crude Odds Ratio (95% CI)	p-value
		99 (51.8)	**92 (48.2)**		
Gum Disease					
No	151 (79.1)	90 (90.9)	61 (66.3)		
Yes	40 (20.9)	9 (9.1)	31 (33.7)	5.08 (2.26, 11.43)	**0.001**
Gum Health					
Healthy	131 (68.6)	79 (79.8)	52 (56.5)		**0.001**
Not Healthy	60 (31.4)	20 (20.2)	40 (43.5)	3.04 (1.60, 5.77)	
Gum Treatment History					
No	77 (40.3)	38 (38.4)	39 (42.4)		
Yes	114 (59.7)	61 (61.6)	53 (57.6)	0.85 (0.48, 1.51)	0.573
Loose Teeth					
No	156 (81.7)	91 (91.9)	65 (70.7)		
Yes	35 (18.3)	8 (8.1)	27 (29.3)	4.73 (2.02, 11.06)	**0.001**
Bone Loss					
No	168 (88.0)	96 (97.0)	72 (78.3)		
Yes	23 (12.0)	3 (3.0)	20 (21.7)	8.89 (2.54, 31.07)	**0.001**
Tooth Position Change					
No	169 (88.5)	91 (91.9)	78 (84.8)		
Yes	22 (11.5)	8 (8.1)	14 (15.2)	0.04 (0.81, 5.12)	0.123

Logistic Regression: Bold indicate significance at p-value < 0.05.

Multivariable Models and Internal Validation

Three SROH variables (self-reported gum disease, loose teeth and alveolar bone loss) remained independently associated with periodontitis in the multivariable analysis (Table 3). In the combined model, age was the only additional predictor retained. Participants aged < 40 years had 73% lower odds of having periodontitis than those ≥ 40 years (aOR = 0.27, 95% CI: 0.11–0.62; p = 0.002).

Table 3
Multivariable Logistic Regression and Diagnostic Accuracy Metrics

Predictors Included in Model	Regression Coefficient (β)	Standard Error (SE)	Adjusted Odds Ratio (95% CI)	p-value
Model 1: SROH Only				
Intercept	-0.478	0.265	0.62 (0.37, 1.04)	0.071
Gum_Disease				
No				
Yes	1.262	0.502	3.53 (1.35, 9.83)	**0.012**
Loose_Teeth				
No				
Yes	1.339	0.679	3.81 (1.43, 11.19)	**0.010**
Bone_Loss				
No				
Yes	1.476	0.679	4.38 (1.29, 20.15)	**0.030**
Model 2: SROH + Risk Factors				
Intercept	1.322	1.008	3.75 (0.55, 33.56)	0.189
Gum_Disease				
No				
Yes	1.282	0.518	3.61 (1.33, 10.34)	**0.013**
Loose_Teeth				
No				
Yes	1.237	0.537	3.44 (1.24,10.41)	**0.021**
Bone Loss				
No				
Yes	1.445	0.696	4.24 (1.19, 20.02)	**0.038**
Age Category				
Below 40	-1.301	0.428	0.27 (0.11, 0.62)	**0.002**
40 And Above				

Logistic Regression: Bold indicate significance at p-value < 0.05.

The apparent AUROC was 0.714 for the SROH-only model and 0.763 for the combined model. Following bootstrap internal validation, the optimism-corrected AUROC values were 0.702 and 0.745, respectively (Table 4). At the model-specific thresholds determined using the Youden index, the addition of age also increased sensitivity from 52.2% to 70.7% but reduced specificity from 83.8% to 74.7%. The combined model had a positive predictive value of 72.2% and a negative predictive value of 73.3%.

Table 4
Model Performance

Diagnostic Performance Evaluation	**Model 1 Metrics**	**Bootstrap 95% CI***	**Model 2 Metrics**	**Bootstrap 95% CI***
Apparent AUROC	0.714	(0.64, 0.78)	0.763	(0.69, 0.83)
Optimism-Corrected AUROC	0.702	(0.00, 1.00)	0.745	(0.00, 1.00)
Sensitivity (%)	52.20	(42.40, 63.00)	70.70	(59.80, 79.30)
Specificity (%)	83.80	(76.80, 90.90)	74.70	(66.70, 82.80)
Positive Predictive Value (PPV)	75.00	NA	72.20	NA
Negative Predictive Value (NPV)	65.40	NA	73.30	NA
***Confidence Interval**				

Model Discrimination and Specification

The combined model showed greater discriminant than the SROH-only model (Fig. 1). The apparent AUROC increased from 0.714 for the SROH-only model to 0.763 for the combined model. Following bootstrap internal validation, the optimism-corrected AUROC values were 0.702 and 0.745, respectively.

Fig. 1 - Receiver Operating Characteristics (ROC) Curves

The final combined prediction model (Model 2) was expressed as:

logit(P) = 1.322 + 1.282(GumDisease) + 1.237(LooseTeeth) + 1.445(BoneLoss) − 1.301(Age < 40)

where *P* represents the probability of periodontitis. Self-reported gum disease, loose teeth, and alveolar bone loss were coded as 1 for “Yes” and 0 for “No”. Age < 40 years was coded as 1, and age ≥ 40 years was coded as 0. The predicted probability of having periodontitis was calculated using:

$$P = \frac{e^{\text{logit}(P)}}{1 + e^{\text{logit}(P)}}$$

The bootstrap-corrected calibration curve showed generally close agreement between predicted and observed probabilities of periodontitis (Fig. 2). Some deviation from the ideal line was observed, with a tendency to underestimate risk at lower predicted probabilities and overestimate risk at higher predicted probabilities.

Decision curve analysis demonstrated that the combined model provided a higher net benefit than referring all participants or referring none across a range of threshold probabilities (Fig. 3). At lower thresholds, the net benefit of the model was similar to the strategy of referring all participants.

The final combined prediction model was translated into a nomogram self-reported gum disease, loose teeth, alveolar bone loss, and age to facilitate estimation of an probability of having periodontitis (Fig. 4).

Discussion

This study develop and internally validated an SROH-based prediction model for identifying individuals at risk of having periodontitis. The final combined model included self-reported gum disease, loose teeth, alveolar bone loss and age. Compared with the SROH-only model, adding age improved overall discrimination and increased sensitivity, although this was accompanied by lower specificity. The combined model demonstrated acceptable optimism-corrected discrimination, generally satisfactory calibration and potential clinical usefulness across a range of threshold probabilities.

The small difference between the apparent AUROC of 0.763 and the optimism-corrected AUROC of 0.745 suggests limited optimism in the model's estimated discrimination. Bootstrap resampling enabled the full study sample to be used for model development while providing a more realistic estimate of expected performance than the apparent AUROC alone [24–27]. Nevertheless, internal validation does not establish the model's transportability, and external validation in an independent population is required before it can be recommended for routine community screening.

The three SROH predictors retained in the final model were clinically plausible indicators of previous periodontal tissue destruction [11, 28–31]. Self-reported alveolar bone loss made the largest contribution to the model. This may be because alveolar bone loss represents irreversible cumulative periodontal damage and is usually communicated to patients following professional assessment. Individuals who have previously been told that they have bone loss may therefore recall this information more reliably than less specific or transient periodontal symptoms [11, 32–33] .

Self-reported gum disease and loose teeth also remained independently associated with periodontitis. Tooth mobility is recognisable manifestation of periodontal breakdown and be more easily perceived by individuals than less apparent clinical changes, although it is more likely to reflect established rather than early disease [11, 29–30]. Previous studies have similarly identified self-reported gum disease, tooth mobility and bone loss as useful components of questionnaire-based periodontal screening models [6, 10]. The present findings support the use of these self-reported indicators to identify individuals who may benefit from comprehensive periodontal assessment.

The findings are consistent with studies conducted in the European populations. Saka-Herrán et al. [34] found that combinations of self-reported periodontal indicators, particularly tooth mobility and gingival changes, were useful for periodontitis screening in a Spanish population. Similarly, Machado et al. [11] reported that self-reported periodontal history combined with indicators of cumulative tissue destruction demonstrated good predictive performance in a Portuguese population. These findings suggest that individuals can recognize and recall meaningful manifestations of periodontal tissue destruction, supporting the use of self-reported indicators as a practical first-line approach for identifying those who may require full periodontal assessment.

Age provided additional predictive information beyond the SROH indicators. Although age is not a direct cause of periodontitis, it may reflect cumulative exposures to behavioural, environmental, and systemic factors associated with periodontal tissue destruction [28, 35–36]. Participants aged ≥ 40 years had higher odds of having periodontitis than younger participants. Adding age increased the model's sensitivity from 52.2% to 70.0%, although specificity decreased from 83.8% to 74.7%. This trade-off may be acceptable for a community triage tool intended to identify individuals who should be referred for further assessment, but the appropriate operating threshold should be evaluated during external validation. Age is also readily available and does not require clinical examination, allowing it to improve risk estimation without substantially increasing the complexity of the model.

The model may also have potential applications beyond conventional dental settings. Its four predictors can be obtained without clinical examination, specialised equipment or laboratory testing, making it potentially suitable for community health programs, primary healthcare clinics, and chronic disease services. Previous studies have similarly explored questionnaire-based periodontal screening in medical settings. Verhulst et al. [8] demonstrated the potential of combining self-reported periodontal information with demographic and behavioural characteristics for opportunistic screening, while Nijland et al. [7] showed that model performance may vary when existing tool is applied to a different medical population. These findings support the potential role of SROH-based models outside dental care while reinforcing the need for external validation in the intended setting. Following validation, the present model could potentially be incorporated into web-based questionnaires, mobile applications or electronic health records to facilitate large-scale periodontal risk assessment and referral.

Clinical Implications

The final model uses four readily obtainable variables and can be administered without specialised equipment or clinical examinations. Its intended role is to alert individuals who may be at risk of having periodontitis and encourage them to seek comprehensive periodontal assessment. It may therefore support initial risk assessment and referral in community and non-dental settings where routine periodontal examination is not readily available. However, the model is not intended to confirm or exclude periodontitis and should not replace a comprehensive periodontal examination. External validation and evaluation of an appropriate referral threshold are required before routine implementation.

Limitations

Several limitations should be considered. First, the SROH questionnaire and periodontal examination were not conducted concurrently. However, the outcome was the presence or absence of periodontitis rather than current disease activity or stage. Periodontitis represents a history of irreversible periodontal attachment loss, while most SROH items captured previous diagnoses or experiences. The interval was therefore unlikely to materially alter the binary classification of periodontitis status. Nevertheless, concurrent collection of self-reported and clinical data should be considered in future validation studies.

Second, self-reported information is subject to recall and reporting biases and depends on individuals' awareness of their periodontal status. Questions concerning previous diagnoses of gum disease or alveolar bone loss may perform better among individuals who have previously received dental care. Conversely, people with undiagnosed or less apparent disease may not report these indicators, particularly because early periodontitis may be asymptomatic. The model may therefore be more effective in identifying established periodontitis than less apparent disease, and its performance may vary according to oral health literacy, previous dental attendance and access to dental care [6, 10–11, 37].

Finally, the study was conducted at a single institution using an available sample of 191 participants, and only 63.7% of the eligible parent-study participants took part. The study population was predominantly female, tertiary educated and had never smoked, which may limit generalisability to the wider community. The modest sample size and data-driven predictor selection may also have affected model stability. Although bootstrap validation was used to estimate and correct for optimism, the model requires external validation in larger and more diverse populations. The predictive values and optimal referral threshold should also be reassessed in populations with different periodontitis prevalence.

Future Research

The model should be externally validated in larger and more diverse community and primary healthcare populations, including individuals with different levels of previous dental attendance and awareness of their periodontal status. Future studies should collect SROH and clinical examination data concurrently and assess whether recalibration or modification of the referral threshold is required in populations with different periodontitis prevalence. Following external validation, implementation studies should evaluate the model's feasibility, acceptability, digital integration, effect on referral pathways and cost-effectiveness.

Conclusion

This study developed and internally validated an SROH-based prediction model incorporating self-reported gum disease, loose teeth, alveolar bone loss and age. The combined model demonstrated acceptable discrimination, generally satisfactory calibration and potential clinical utility for identifying individuals who may be at risk of having periodontitis. It may serve as a simple, non-invasive triage tool to alert individuals and support referral for full periodontal examination in community and non-dental

settings. External validation is required before the model can be recommended for routine implementation.

Abbreviations

AAP
American Academy of Periodontology
AUROC
Area under the receiver operating characteristic curve
CAL
Clinical attachment loss
CI
Confidence interval
DCA
Decision curve analysis
EFP
European Federation of Periodontology
NPV
Negative predictive value
PPD
Probing pocket depth
PPV
Positive predictive value
SROH
Self-reported oral health
TRIPOD
Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis

Declarations

Acknowledgements

The authors also thank the Faculty of Dentistry Medical Ethics Committee, Universiti Malaya, for its ethical review and approval, and all participants for their contribution to this study.

Authors' contributions

Concept and design: Nor Adinar Baharuddin, Ainol Haniza Kherul Anuwar, Alias Abd Aziz. Acquisition, analysis, or interpretation of data: Tayf Naif R Alshammari, Nor Adinar Baharuddin, Ainol Haniza Kherul Anuwar, Alias Abd Aziz. Initial draft: Alias Abd Aziz. Revision and editing: Nor Adinar Baharuddin, Ainol Haniza Kherul Anuwar, Alias Abd Aziz. Resources: Nor Adinar Baharuddin, Ainol Haniza Kherul Anuwar, Alias Abd Aziz. Approval of manuscript: All authors.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available due to ethical restrictions related to participant confidentiality and institutional approval requirement. Anonymised data may, however, be made available by the corresponding author upon reasonable request, subject to institutional approval and compliance with applicable ethical and data protection requirements.

Ethics Statement and consent to participate

The parent study was approved by the Faculty of Dentistry Medical Ethics Committee, Universiti Malaya (Reference No. DF RD2401/0010 (L)). The present study, which involved administration of the self-reported questionnaire and linkage with existing periodontal examination data, received separate ethical approval from the same committee (Reference No. DF OC2503/0073 (P)). Written informed consent was obtained from all participants.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Authors details

[1]Department of Oral and Craniofacial Sciences, Faculty of Dentistry, Universiti Malaya, Kuala Lumpur, Malaysia.

[2]Department of Restorative Dentistry, Faculty of Dentistry, Universiti Malaya, Kuala Lumpur, Malaysia.

[3]Department of Community Oral Health & Clinical Prevention, Faculty of Dentistry, Universiti Malaya, Kuala Lumpur, Malaysia.

[4]Oral Cancer Research & Coordinating Centre (OCRCC), Faculty of Dentistry, Universiti Malaya, 50603 Kuala Lumpur, Malaysia.

[5]Adjunct Professor, Faculty of Dental Medicine Universitas Airlangga, Surabaya, Indonesia.

References

1. Kassebaum N, Bernabé E, Dahiya M, et al. Global burden of severe periodontitis in 1990-2010: a systematic review and meta-regression. J Dent Res 2014;93(11):1045-1053.
2. Nazir MA. Prevalence of periodontal disease, its association with systemic diseases and prevention. Int J Health Sci (Qassim) 2017;11(2):72.
3. World Health Organization. Global oral health status report: towards universal health coverage for oral health by 2030. Geneva: World Health Organization; 2022.
4. Ministry of Health Malaysia. Standard operating procedure for Ministry of Health dental clinics: management of periodontal conditions in primary oral healthcare clinics. Kuala Lumpur: Ministry of Health Malaysia, Oral Health Programme; 2018.
5. Barbosa C, Manso MC, Reis T, et al. Are oral overuse behaviours associated with painful temporomandibular disorders? A cross-sectional study in Portuguese university students. J Oral Rehabil 2021;48(10):1099-1108. doi:10.1111/joor.13226.
6. Iwasaki M, Usui M, Ariyoshi W, et al. Validation of a self-report questionnaire for periodontitis in a Japanese population. Sci Rep 2021;11(1):15078.
7. Nijland N, Su N, Gerdes V, et al. Attempts to modify periodontal screening models based on a self-reported oral health questionnaire in the medical care setting. J Clin Periodontol 2025;52(3):387-398.
8. Verhulst MJ, Teeuw WJ, Bizzarro S, et al. A rapid, non-invasive tool for periodontitis screening in a medical care setting. BMC Oral Health 2019;19(1):87.
9. Deng K, Pelekos G, Jin L, et al. Diagnostic accuracy of self-reported measures of periodontal disease: a clinical validation study using the 2017 case definitions. J Clin Periodontol 2021;48(8):1037-1050.
10. Eke P, Dye B, Wei L, et al. Self-reported measures for surveillance of periodontitis. J Dent Res 2013;92(11):1041-1047.
11. Machado V, Lyra P, Santos C, et al. Self-reported measures of periodontitis in a Portuguese population: a validation study. J Pers Med 2022;12(8):1315.
12. Iba K, Shinozaki T, Maruo K, et al. Re-evaluation of the comparative effectiveness of bootstrap-based optimism correction methods in the development of multivariable clinical prediction models. BMC Med Res Methodol 2021;21(1):9.
13. Moons KG, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162(1):W1-W73.
14. Riley RD, Collins GS. Stability of clinical prediction models developed using statistical or machine learning methods. Biometr J 2023;65(8):2200302.
15. Ho JY, Anuwar AHK, Karen-Ng LP, et al. Non-invasive periodontal screening using self-reported-oral-health (SROH) questionnaire and salivary biomarkers: development and validation of machine learning models. BMC Oral Health 2026. doi:10.1186/s12903-026-08759-9.

16. Bi M, Xie Y, Yu X, et al. Diagnostic accuracy of self-reported questionnaires for detecting periodontitis across multiple cultures and geographic locations: a systematic review and meta-analysis. J Clin Periodontol 2025;52(11):1510-1528.
17. Sim CP, Li H, Peres MA. Who to be treated: nomogram using self-reported periodontal screening instrument among English-speaking adults in multi-ethnic Singapore. J Pers Med 2022;12(6):931.
18. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Circulation 2015;131(2):211-219. doi:10.1161/CIRCULATIONAHA.114.014508.
19. Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: part II—binary and time-to-event outcomes. Stat Med 2019;38(7):1276-1296.
20. Caton JG, Armitage G, Berglundh T, et al. A new classification scheme for periodontal and peri-implant diseases and conditions: introduction and key changes from the 1999 classification. J Periodontol 2018;89(Suppl 1):S1-S8. doi:10.1002/JPER.18-0157.
21. Cai X, Yang J, Luo S, et al. Development and internal validation of a nomogram for predicting referral or hospitalization risk in children with community-acquired influenza. Front Pediatr 2026;14:1779925. doi:10.3389/fped.2026.1779925.
22. Li C, Guan J, Zhao Q, et al. A visualized nomogram to predict the risk of acute ischemic stroke among patients with cervical artery dissection. Int J Gen Med 2025;18:1569-1580. doi:10.2147/IJGM.S507043.
23. Ying L, Yangmei S, Qing X, et al. Construction and validation of a risk prediction model for symptom cluster distress in brain tumor patients: a cross-sectional study. BMC Cancer 2025;25(1):1803. doi:10.1186/s12885-025-15281-8.
24. Cai B, Luo Y, Guo X, et al. Bootstrapping the cross-validation estimate. Ann Appl Stat 2025;19(4):2981-3002.
25. Collins GS, Dhiman P, Ma J, et al. Evaluation of clinical prediction models (part 1): from development to external validation. BMJ 2024;384:e074819. doi:10.1136/bmj-2023-074819.
26. Fernandez-Felix BM, García-Esquinas E, Muriel A, et al. Bootstrap internal validation command for predictive logistic regression models. Stata J 2021;21(2):498-509. doi:10.1177/1536867X211025836.
27. Zhang C, Yan R, Liu X, et al. Empirical simulation of internal validation methods for prediction models: comparing k-fold cross-validation with bootstrap-based optimism correction. J Clin Epidemiol 2026;190. doi:10.1016/j.jclinepi.2025.112101.
28. Carra MC, Gueguen A, Thomas F, et al. Self-report assessment of severe periodontitis: periodontal screening score development. J Clin Periodontol 2018;45(7):818-831. doi:10.1111/jcpe.12899.
29. Heaton B, Gordon NB, Garcia RI, et al. A clinical validation of self-reported periodontitis among participants in the Black Women's Health Study. J Periodontol 2017;88(6):582-592.
30. Khader Y, Alhabashneh R, Alhersh F. Development and validation of a self-reported periodontal disease measure among Jordanians. Int Dent J 2015;65(4):203-210.

31. Krois J, Graetz C, Holtfreter B, et al. Evaluating modeling and validation strategies for tooth loss. J Dent Res 2019;98(10):1088-1095.
32. Mendes JJ, Neves M, Supiot C, et al. Combining self-reported information with radiographic bone loss to screen periodontitis: a performance study. J Clin Med 2025;14(13):4531.
33. Vishwakarma P, Ray PM, Dodamani A, et al. Comparative assessment of modified self-reported periodontal risk assessment model and periodontal risk assessment model among the adult population. Indian J Community Med 2023;48(1):70-74.
34. Saka-Herrán C, Jané-Salas E, González-Navarro B, et al. Validity of a self-reported questionnaire for periodontitis in a Spanish population. J Periodontol 2020;91(8):1027-1038. doi:10.1002/JPER.19-0604.
35. Chatzopoulos GS, Jiang Z, Marka N, et al. Periodontal disease, tooth loss, and systemic conditions: an exploratory study. Int Dent J 2024;74(2):207-215. doi:10.1016/j.identj.2023.08.002.
36. LaMonte MJ, Hovey KM, Millen AE, et al. Accuracy of self-reported periodontal disease in the Women's Health Initiative Observational Study. J Periodontol 2014;85(8):1006-1018. doi:10.1902/jop.2013.130488.
37. Lertpimonchai A, Tuntrakul S, Rattanasiri S, et al. Validity of simple self-reported periodontal status questions. Int Dent J 2023;73(1):121-127. doi:10.1016/j.identj.2022.10.006.

Figures

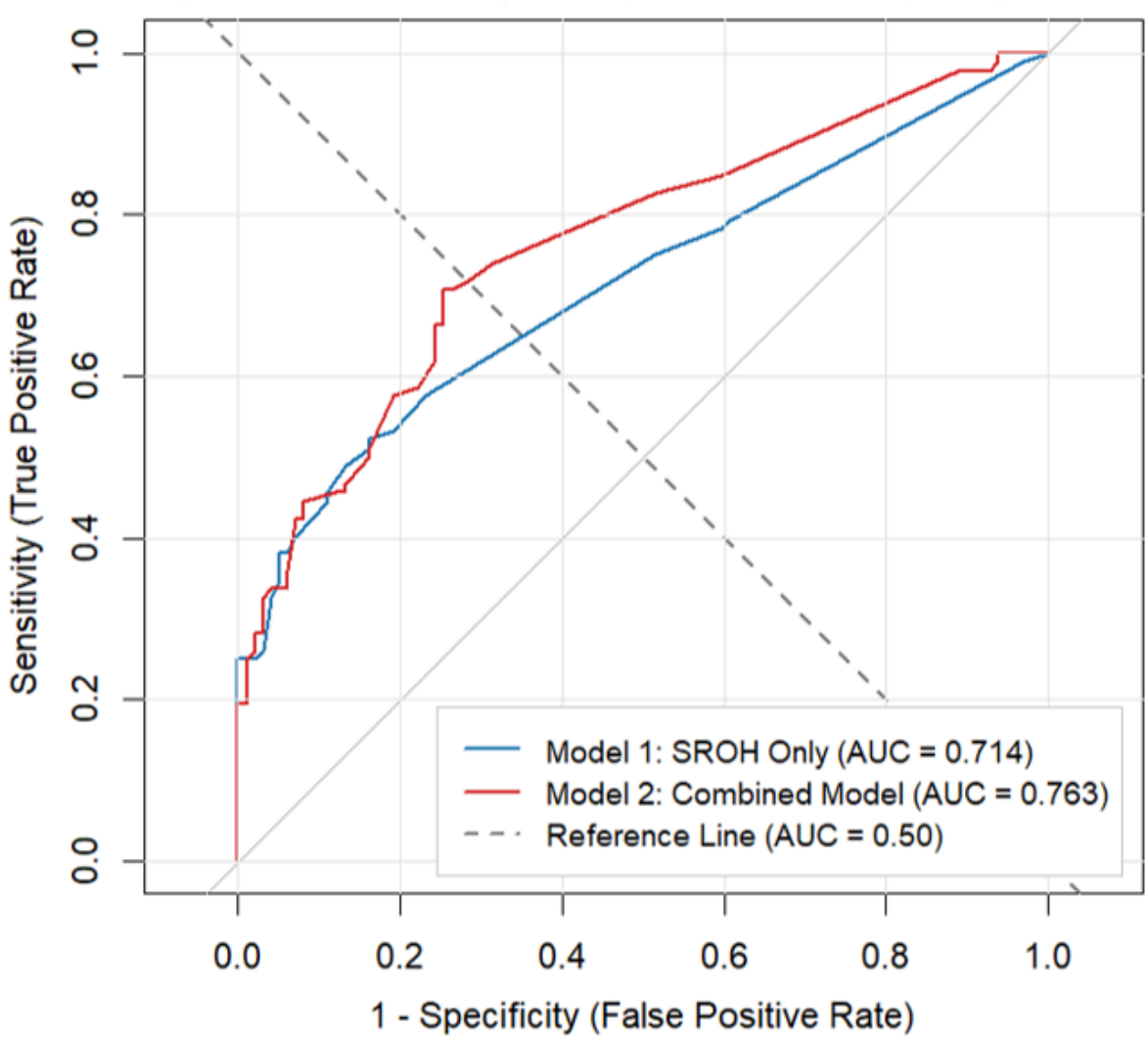


Figure 1

Receiver Operating Characteristics (ROC) Curves

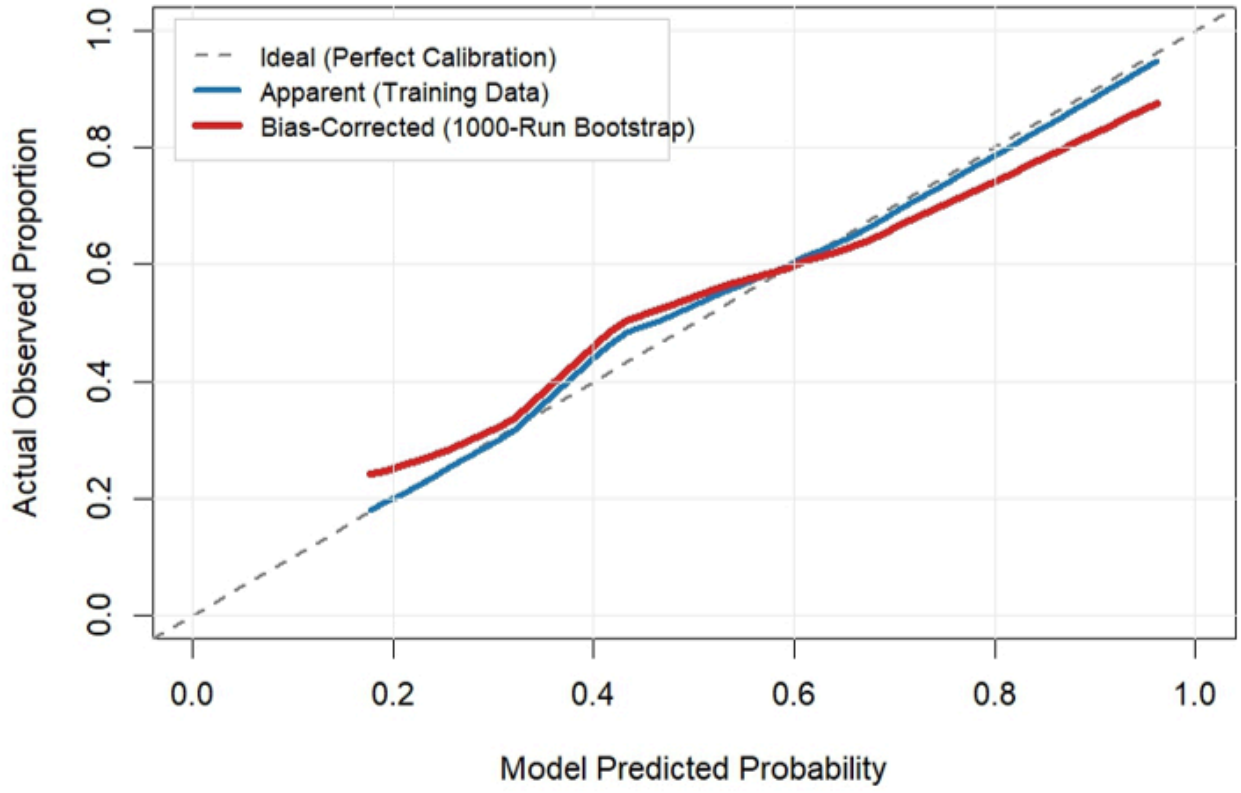


Figure 2

Optimism-Corrected Calibration Curve (Model 2)

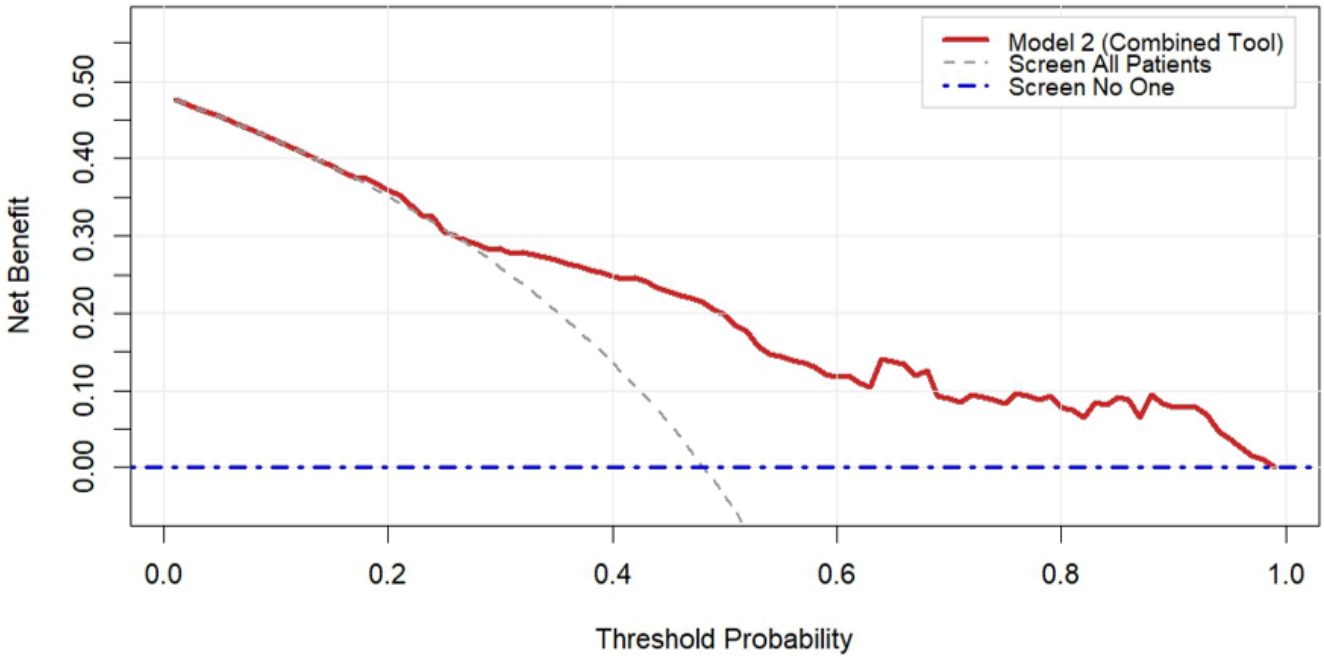


Figure 3

Decision Curve Analysis (Clinical Net Benefit)

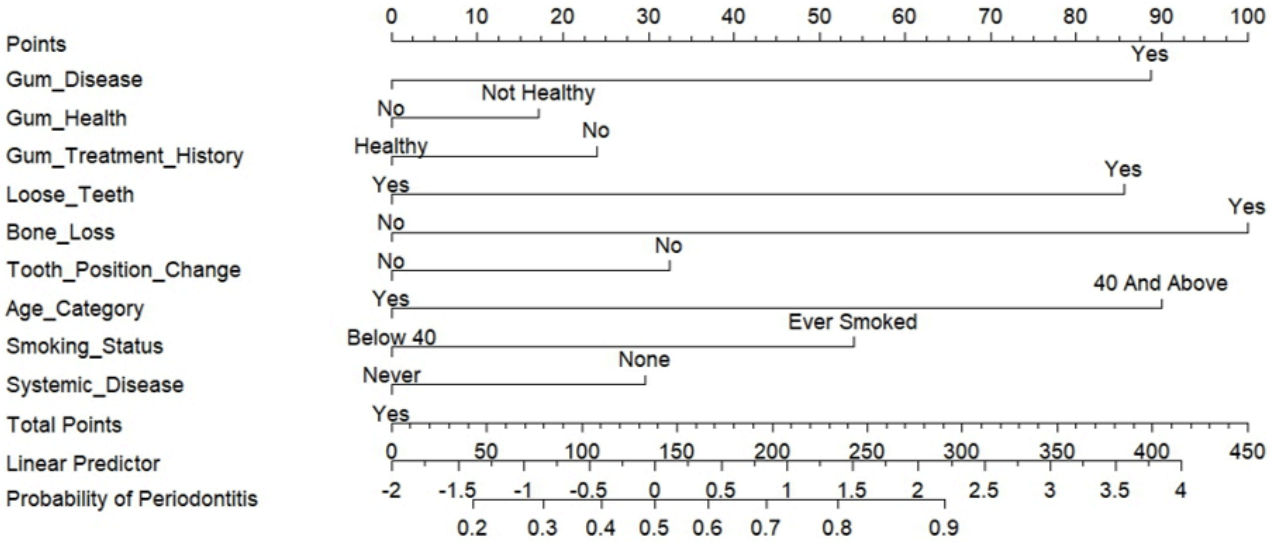


Figure 4

Diagnostic Nomogram Scoring Predictive Architecture

Supplementary Files

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

- Suplemenatary1QuestionnaireMSROHcopy.pdf
- Supplementary2TRIPODchecklist.docx