

Received: 4 September 2020

Revised: 26 February 2021

Accepted: 28 February 2021

DOI: 10.1111/jcpe.13455

ORIGINAL ARTICLE CLINICAL PERIODONTOLOGY

 WILEY

Have you had bleeding from your gums? Self-report to identify giNGival inflammation (The SING diagnostic accuracy and diagnostic model development study)

Beatriz Goulão  | Graeme S. MacLennan | Craig R. Ramsay

Health Services Research Unit, Centre for Healthcare Randomized Trials, University of Aberdeen, Aberdeen, UK

Correspondence

Beatriz Goulão, Health Services Research Unit, Health Sciences Building, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, UK.

Email: beatriz.goulao@abdn.ac.uk

Funding information

IQuaD was funded by the NIHR HTA (project 09/01/45). INTERVAL was funded by the NIHR HTA (project 06/35/99). No additional funding was obtained to conduct the diagnostic accuracy and prediction studies presented in this paper. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The Health Services Research Unit is funded by the Chief Scientist Office of the Scottish Government Health and Social Care Directorates.

Abstract

Aim: To assess the diagnostic performance of self-reported oral health questions and develop a diagnostic model with additional risk factors to predict clinical gingival inflammation in systemically healthy adults in the United Kingdom.

Methods: Gingival inflammation was measured by trained staff and defined as bleeding on probing (present if bleeding sites \geq 30%). Sensitivity and specificity of self-reported questions were calculated; a diagnostic model to predict gingival inflammation was developed and its performance (calibration and discrimination) assessed.

Results: We included 2853 participants. Self-reported questions about bleeding gums had the best performance: the highest sensitivity was 0.73 (95% CI 0.70, 0.75) for a Likert item and the highest specificity 0.89 (95% CI 0.87, 0.90) for a binary question. The final diagnostic model included self-reported bleeding, oral health behaviour, smoking status, previous scale and polish received. Its area under the curve was 0.65 (95% CI 0.63–0.67).

Conclusion: This is the largest assessment of diagnostic performance of self-reported oral health questions and the first diagnostic model developed to diagnose gingival inflammation. A self-reported bleeding question or our model could be used to rule in gingival inflammation since they showed good sensitivity, but are limited in identifying healthy individuals and should be externally validated.

KEYWORDS

diagnosis, epidemiology, gingival inflammation, prediction modelling, self-report

Clinical Relevance

Scientific rationale for Study: Gingival inflammation is a precursor of periodontitis and needs to be identified in population-based studies; however, it is costly and difficult to measure. Self-reported questions can be an efficient and inexpensive alternative to clinically assessed measures in epidemiological surveillance.

Principal Findings: A self-reported bleeding gums question had high sensitivity and low specificity to identify clinically measured gingival inflammation. Our diagnostic prediction model for gingival inflammation had moderate discriminatory ability.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Journal of Clinical Periodontology* published by John Wiley & Sons Ltd.

Practical Implications: A self-reported bleeding gums question or a diagnostic model is helpful to identify people with gingival inflammation and should be considered if unfeasible or impossible to collect clinical data.

1 | INTRODUCTION

Periodontal disease is an inflammatory disease that affects the soft and hard tissues supporting teeth. This disease is largely preventable, yet it remains the major cause of poor oral health worldwide and is the primary cause of tooth loss in older adults (Petersen & Ogawa, 2000). Periodontal disease is classified into two broad categories: (Petersen & Ogawa, 2000) gingivitis and (Tonetti et al., 2020) periodontitis. Gingivitis is usually characterized by, amongst other factors, gingival inflammation, which is a reversible condition identified by bleeding at the gingival margin. Gingival inflammation is an early sign of periodontal disease, (Tonetti et al., 2020) and its association with periodontal health has been explored in numerous publications (Kallio, 1996; Weintraub et al., 2013). Identifying gingival inflammation could help prevent periodontal disease.

The development of self-reported tools to measure periodontal disease risk is particularly important in the field of oral health surveillance (Ramos et al., 2013), since clinical measures are difficult and costly to collect, and hard or impossible to standardize. The Centres for Disease Control and Prevention (CDC) and the American Academy of Periodontology recommended since 2003 the use of self-reported measurements that could be valid to predict the prevalence of periodontal disease as an alternative to examinations (Eke et al., 2012). Several self-reported oral health questions have been previously discussed as an alternative to clinical measures of periodontal disease (Blicher et al., 2015; Abbood et al., 2016), but studies focusing on gingivitis or gingival inflammation have been inconclusive due to small, non-representative samples and the use of single questions without the consideration of other potential risk factors (Abbood et al., 2016).

Prediction models and specifically diagnostic models allow the inclusion of several risk factors to predict the existence of a condition (Collins et al., 2015). Prediction models are currently used in different areas of medicine (Collins et al., 2015) and have been increasingly popular in assessing the prevalence and progression of periodontitis (Du et al., 2018). There are currently no models available to predict gingival inflammation. Diagnostic models of gingival inflammation could be used as a first line assessment in a clinical surveillance system, identifying patients that need further assessment, or they could be incorporated in large epidemiological studies to target clinical examinations in research participants.

We aimed to assess the diagnostic accuracy of several self-reported oral health questions individually to identify gingival inflammation, measured by bleeding on probing, in a large and representative UK-based sample. We developed and validated a diagnostic model including other gingival inflammation risk factors to assess whether additional risk factors would improve the

diagnostic accuracy of these measures. We hypothesized that using self-reported oral health questions with or without additional risk factors could help identify clinical gingival inflammation and, therefore, constitute an alternative to clinical measurement in research settings.

2 | METHODS

2.1 | Data

We collected and combined clinical and self-reported data from adults participating in two large, UK-wide dental randomized trials. Combination of data from both trials was deemed reasonable to follow TRIPOD guidelines and maximize statistical precision (Collins et al., 2015) and because inclusion criteria, setting, recruitment and data collection processes, patient reported and clinical outcomes collected were identical in both trials. Moreover, the trials were conducted by the same team of researchers.

2.1.1 | IQuaD trial

The IQuaD trial (Ramsay et al., 2018) compared the clinical and cost-effectiveness of providing scale and polish 6 monthly, 12 monthly or never during 3 years; the clinical and cost-effectiveness of personalized vs standard oral hygiene advice were also compared. The study recruited 1877 participants from Scotland and Northeast England. Recruitment began in 2011, and data were collected until September 2016.

2.1.2 | INTERVAL trial

The INTERVAL trial (Clarkson et al., 2018) compared the clinical and cost-effectiveness of 24-monthly recalls vs 6-monthly vs risk-based on the same outcome and randomized 2372 participants from Scotland, England, Wales and Northern Ireland. Recruitment began in July 2010, and data were collected until August 2018.

Both trials were pre-registered and had bleeding on probing as their primary outcome. They were set up in primary dental care provided by the National Health System in the United Kingdom. Participants were recruited via their dental practices.

Participant-reported data, including the index tests, in IQuaD and INTERVAL were collected using an annual patient questionnaire from baseline until year 3 (for IQuaD) or 4 (for INTERVAL). The data used in our model are restricted to the final year follow-up of

TABLE 1 Four self-reported gingivitis measures used and their scores and transformation for individual diagnostic analysis

Self-reported measures	Original score (used in the diagnostic model)	Transformation for individual diagnostic accuracy analysis
Self-report bleeding gums (Likert) (<i>'Have you had bleeding from your gums when brushing your teeth?'</i>)	1:Never; 2:Hardly ever; 3:Occasionally; 4:Fairly often; 5:Very often.	A score of 2 or more was considered as a positive index test
Self-report bleeding gums (Binary) (<i>'Do your gums bleed when brushing?'</i>)	Yes/No If participants replied, 'not sure', they were deemed as missing	Yes/No If participants replied, 'not sure', they were deemed as missing
Self-report unpleasant taste in your mouth (<i>'Have you had an unpleasant taste in your mouth when brushing your teeth?'</i>)	1:Never; 2:Hardly ever; 3:Occasionally; 4:Fairly often; 5:Very often.	A score of 2 or more was considered as a positive index test
Self-report bad breath (<i>'Have you had bad breath?'</i>)	1:Never; 2:Hardly ever; 3:Occasionally; 4:Fairly often; 5:Very often.	A score of 2 or more was considered as a positive index test

each trial; therefore, this is a cross-sectional analysis. IQuaD and INTERVAL collected similar information about the participants, but not all outcomes were collected in both trials.

Clinical data were collected at follow-up (3 or 4 years post-randomization for IQuaD and INTERVAL, respectively). Variables collected at the clinical assessment included the following: bleeding on probing, calculus and pocket depth. Clinical outcomes were measured at the end of each trial by trained outcome assessors. Gingival inflammation was measured according to the Gingival Index of Loe (1967) by running the UNC probe circumferentially around each tooth just within the gingival sulcus or pocket. After 30 s, bleeding was recorded as being present or absent on the buccal and lingual surfaces. Percentage of sites bleeding on probing per participant was calculated by adding the sites where bleeding was present in each participant (two sites per tooth – buccal and lingual) and dividing by twice the number of teeth in the mouth and then multiplied by 100, thus generating a variable varying from 0 (no bleeding in any of the sites available) to 100 (bleeding in all sites available). More information about clinical outcome collection can be found in IQuaD's and INTERVAL's protocols (Clarkson et al., 2013; Clarkson et al., 2018). Participants did not have access to their clinical data before answering the questions and clinical examiners did not have access to self-reported data.

2.2 | Study population

Eligibility criteria were similar in both trials: both included adult participants (≥ 18 years of age) who were dentate, had attended a check-up at least twice in the 2 years prior to the trial and received their dental care in part or fully as a patient in the National Health System in the United Kingdom. Patients with an uncontrolled medical condition (e.g. diabetes and immunocompromised) were excluded from both trials. IQuaD had an additional inclusion criterion to INTERVAL: participants had to score 0–3 in their Basic Periodontal Examination. Participants were approached and recruited if they met the eligibility criteria and accepted to take part; therefore, we used a convenience sample.

2.3 | Gingival inflammation definition

According to the most recent classification from the 2017 Classification of Periodontal Diseases (Chapple et al., 2018; Trombelli et al., 2018), for adults with pocket depths ≤ 3 mm, and an intact and a reduced periodontium, generalized gingivitis for epidemiological purposes can be defined as 30% or more bleeding sites. Localized gingivitis is defined by having between 10 and 29% of bleeding sites. Participants with $< 10\%$ sites bleeding are gingivitis-free. Because pocket depths could be higher than 3 mm and we did not measure clinical attachment level, we did not focus on gingivitis. We used 30% as the primary cut-off for our analyses and considered it to be an indicator of gingival inflammation. To explore the impact of lowering the threshold to define gingival inflammation in our models' secondary analyses used 10% sites bleeding as the cut-off.

2.4 | Diagnostic accuracy of individual self-reported oral health questions

2.4.1 | Outcome

Outcomes are the sensitivity and specificity of the index tests on detection of gingival inflammation (defined as having bleeding on probing equal or higher than 30%). We explored the impact of lowering the threshold to define gingival inflammation (using 10% or greater sites bleeding as the definition) in a sensitivity analysis.

2.4.2 | Index tests

Four self-reported bleeding or bleeding-related measures were used as the index tests in the current study. Table 1 presents the measures and their score. In order to perform the diagnostic accuracy analysis for single self-reported gingival inflammation measures, Likert items were transformed into binary measures. To decide the appropriate threshold, we assessed the scales' diagnostic performance at

different thresholds using receiver operating characteristics curves and selected the best performance. We assumed the best performance to be the one that had the overall best proportion of correctly classified individuals considering both sensitivity and specificity.

2.4.3 | Statistical analysis methods

Diagnostic measures (sensitivity and specificity) were calculated for each self-reported index test measure with 95% confidence intervals calculated using the Agresti-Coull method (Newcombe, 1998). We used a complete case approach where diagnostic measures were calculated if participants had information in both the index test and reference standard (clinical gingival inflammation).

2.5 | Diagnostic model to predict gingival inflammation

2.5.1 | Outcome

The main outcome of interest was generalized gingival inflammation defined as 30% or more sites with bleeding on probing. We explored the impact of lowering the threshold to define gingival inflammation (using 10% or greater sites bleeding as the definition) in a sensitivity analysis.

2.5.2 | Candidate predictors

All variables collected in IQuaD and INTERVAL were considered for inclusion in the model (Table S1). A group of clinical experts, selected by convenience, was approached initially via email and asked to select the most relevant predictors of gingival inflammation. In a face-to-face meeting, the experts were reminded of the survey and there was an opportunity to discuss the topic in smaller groups. Sixteen experts replied to a survey, via Google Docs, available during February 2019. To be considered for inclusion in the model, a predictor had to be selected as important by at least one expert. All predictors selected were included in the 'full model'. The main model excluded clinical predictors (number of teeth, calculus and probing depth). As a sensitivity analysis, we included clinical predictors in the model. For the diagnostic model, self-reported gingival inflammation measures were included as originally scored (as shown in Table 1). Randomized treatment was not included in the model given the cross-sectional nature of the analysis, and the fact that any treatment was only provided to patients after the clinical outcomes was measured.

2.5.3 | Statistical analysis methods

2.6 | Development of a prediction model to identify gingival inflammation

The univariable (unadjusted) associations between each candidate predictor and gingival inflammation were assessed using a logistic

regression model, to assess the impact of each candidate predictor individually in relation to gingival inflammation. We assessed linearity between predictors and the log odds of gingival inflammation and found all relationships to be linear. A p -value of $<.05$ identified statistically significant univariate associations.

2.7 | Candidate predictor selection

All candidate predictors selected by experts were included in an initial 'full model'. To select the final predictors in the final multiple logistic regression model, I implemented an automated backward selection using p -value $<.10$ as the selection criteria. I used this selection threshold because it is a common one and simulation studies show values above 0.05 should be considered (54).

2.8 | Sample size

Sample size was constrained by the number of participants in IQuaD and INTERVAL with a final year clinical assessment. We included participants who had a primary outcome (clinical measure of gingival inflammation), at least one of the self-reported index test measures and predictors collected in both trials.

2.9 | Missing data

We assumed a missing at random mechanism and used multiple imputation as a sensitivity analysis to impute missing data. We excluded predictors from the multiple logistic regression complete case main model that had more than 10% missing data.

2.10 | Performance measures

Performance was evaluated using the full dataset following the appropriate recommendations (Collins et al., 2015). We quantified the ability of the multiple logistic regression model to discriminate between participants with and without gingival inflammation (discriminatory ability) by calculating the area under the receiver operating characteristics curve (c -statistic), sensitivity and specificity (54). The c -statistic gives the probability of a participant with the condition to be given a higher probability of having the condition by the model compared with a randomly chosen participant without the condition (93). Sensitivity and specificity were calculated based on a predicted probability threshold of 0.5. We used this threshold because, at this stage of the model development, we are equally interested in detecting individuals with and without bleeding gums. Calibration of the model (i.e. the agreement between observed gingival inflammation and predicted gingival inflammation) was assessed using a calibration plot (93). The plot assessed this for each 10th

percentile of predicted risk, generating 10 equally sized groups, by plotting observed proportions versus predicted probabilities and adding a smoothed (lowess) line over the entire predicted probability range (93).

2.11 | Internal validation

To assess the model calibration's optimism (defined as the bootstrap performance of the model minus the test performance), we used bootstrapping for internal validation. Two hundred bootstrap samples, each containing the same number of patients, were generated with replacement. The *c*-statistic was calculated for each of the 200 sample-derived models (bootstrap performance) and also calculated for each of the 200 sample-derived models applied to the original dataset (test performance). The difference between the two *c*-statistics for each sample was averaged over the 200 samples. This reflected the optimism of the original model's performance.

Analyses were performed using Stata/SE 15 (StataCorp 2016). A study protocol is available upon request.

2.12 | Ethical considerations

Favourable ethics opinion for the IQuaD trial was confirmed by the East of Scotland Research Ethics Service on 24 March 2011 [Research Ethics Committee (REC) reference number 10/S0501/65]. Favourable ethical opinion for the INTERVAL Dental Recalls Trial was granted by the Fife and Forth Valley Research Ethics Committee [REC Reference number 09/S0501/1]. Written consent was obtained from participants, and more information on the process has been published elsewhere (Ramsay et al., 2018; Clarkson et al., 2018).

3 | RESULTS

We included a maximum of 2853 participants that provided a measure of clinical bleeding and at least one of the index tests (self-reported questions related to gingival inflammation; Figure 1). From those, 1694 (56%) participants had generalized gingival inflammation and 1328 (44%) did not. Overall, 6% of participants with a clinical measure of gingival inflammation were missing all index tests and 80% had all index tests available. The most missed index tests were the questions about unpleasant taste in the mouth when brushing teeth and bad breath with around 15% of participants not providing an answer from those participants that provided a clinical measure. Participants that provided an answer to the index test questions had similar bleeding levels to those that skipped those questions (Table S2). Time between being clinical examined and answering the self-reported questions had a median of 22 days (Percentile 25: 5 days, Percentile 75: 42 days), meaning most participants had their clinical exam after replying to the questionnaire.

3.1 | Summary of candidate predictors

Table 2 shows a summary of candidate predictors. Participants were on average 51 years old, 60% were women, 15% identified as smokers, and 91% identified as regular attenders to the dentist. They had on average 6 out of 9 in an oral health behaviour scale, and they had good quality of life related to oral health (5.2 out 56, where 56 is worst quality of life). Participants had on average 24 teeth (excludes third molars), 36% of sites bleeding, 37% of sites with calculus and a healthy mean pocket depth of 2 mm. Missing data rates for each candidate predictor are presented in the Table S3. We excluded candidate predictors with more than 10% missing data from the main model. For that reason, your last course of treatment was excluded.

3.2 | Diagnostic performance of the self-reported oral health questions

Figures S1-S3 show the ROC curves for each Likert scale question before transformation. The diagnostic performance for the four index tests is given in Table 3 using the main reference standard cut-off of 30%. The binary self-reported bleeding gums question had the highest specificity (0.89) but the lowest sensitivity (0.20). The Likert scale self-reported bleeding gums question had the highest sensitivity (0.73) but poor specificity (0.39). Self-reported unpleasant taste in mouth and bad breath performed poorly. Table S4 shows the diagnostic performance of self-reported oral health questions when using a threshold of 10% with similar results.

3.3 | Diagnostic modelling results

All predictors available were selected at least once by the experts invited and therefore included in the full model. Table S5 shows unadjusted odds ratios with 95% confidence intervals for each predictor. Older age, electric toothbrushes (vs manual), or using both electric and manual toothbrushes (vs manual only) and being a smoker were all significant and associated with lower odds of having gingival inflammation. Regarding self-reported oral health questions, all but bad breath had a significant association with clinical gingival inflammation. Higher oral health behaviour and perceived behaviour control scores were protective of gingival inflammation. Participants with higher calculus levels and deeper pocket depths had higher odds of having gingival inflammation.

In the final model (Table 4), the odds of having gingival inflammation were 1.4 times higher for every additional point in the Likert scale (1- never experienced bleeding from my gums to 5-always experience). Being a smoker and having a higher oral health behaviour score (i.e. doing more appropriate oral health behaviours) were associated with lower odds of having gingival inflammation. Each additional scale and polish in the previous year to participating in the trial were associated with higher odds of having gingival inflammation.

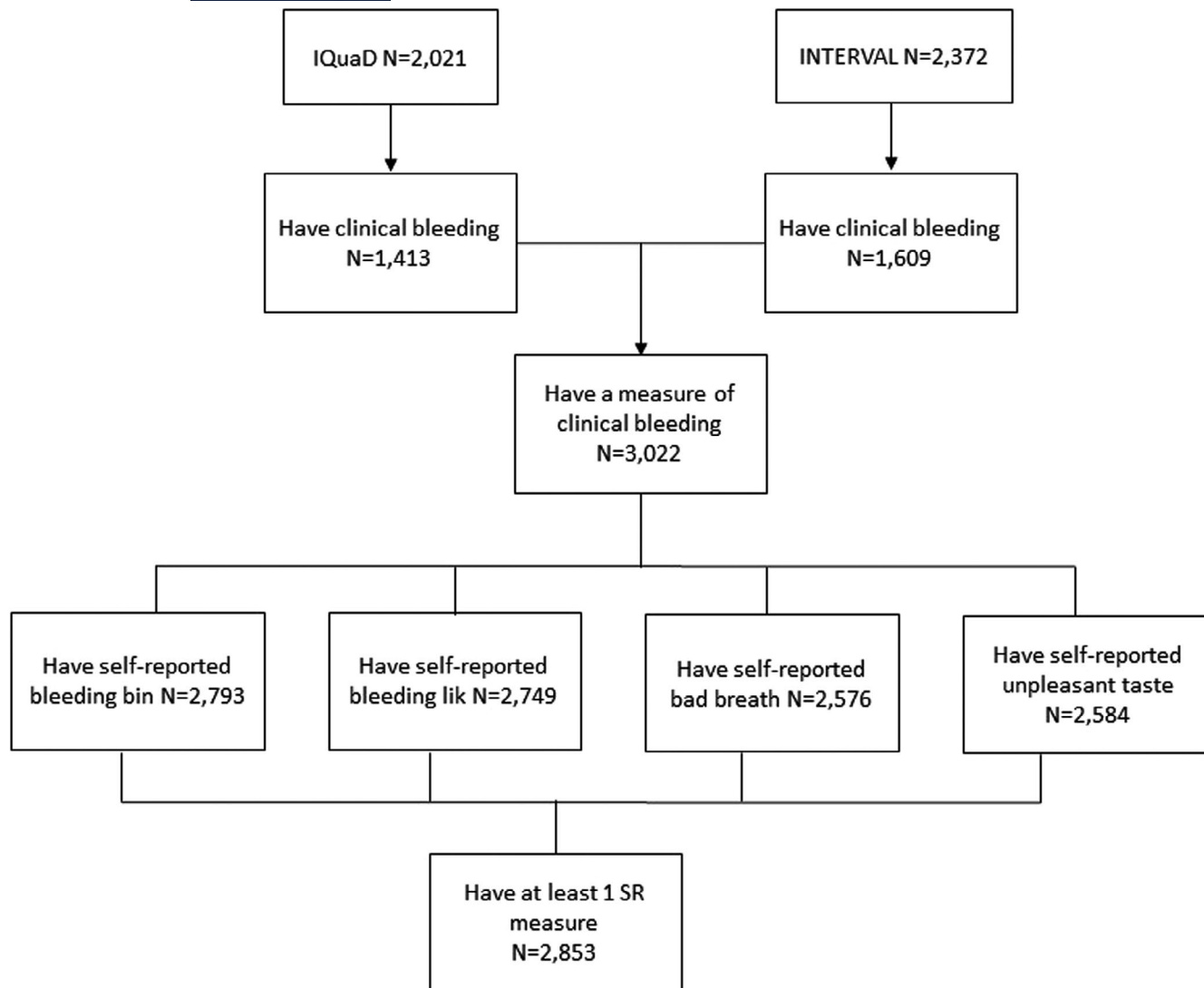


FIGURE 1 Flowchart of the cases available to be analysed in the diagnostic modelling (SR, self-reported; bin, Binary; lik, Likert)

The final model was able to correctly identify participants with gingival inflammation (sensitivity) in 0.74 (95% CI 0.71–0.76) of the cases and able to correctly identify participants without gingival inflammation (specificity) in 0.47 (95% CI 0.44–0.50) of the cases. The trade-offs between sensitivity and specificity for the main model are shown in Figure 2 with the area under the curve. The probability that a patient with gingival inflammation is given a higher probability of having gingival inflammation by the model than a randomly chosen patient without gingival inflammation is 0.65 (c-statistic; 95% CI 0.63–0.67). Optimism in c-statistic was calculated at 0.012 suggesting no model overfitting. A good agreement between observed and predicted probabilities of gingival inflammation was observed in the calibration plot (Figure S4). Using multiple imputation resulted in a similar final model including the same predictors as the main model plus self-reported bleeding gums (binary) with a c-statistic of 0.65. Including clinical predictors resulted in a small increase in the c-statistic (from 0.65 to 0.69) with a similar sensitivity as the model without clinical predictors (0.72), but better specificity (from 0.47 to

0.54). A secondary analysis model using bleeding on probing above 10% as the primary outcome yielded a similar c-statistic of 0.66 (results available upon request). Given the models were based in the aggregation of two studies, IQaD and INTERVAL, we provide the diagnostic performance of the models and self-reported questions for each study in Table S6 and their sensitivity/specificity trade-offs in area under the curve figures (Figures S5–S8).

Figure 3 shows the different discriminatory abilities of self-reported questions and the diagnostic model. The diagnostic model has similar sensitivity to the self-reported bleeding gums question (using a Likert scale) and a higher specificity, but both are poor (below 0.5).

4 | DISCUSSION

To our knowledge, this is the largest study evaluating the diagnostic performance of self-reported oral health questions to detect

TABLE 2 Summary of candidate predictors – mean (SD), count unless indicated otherwise

Candidate predictors	N = 2853
General	
Age	50.6 (14.6), 2853
Female – n (%)	1719 (60.3)
Your last course of treatment was – n (%)	
NHS	2148 (75.3)
Private	126 (4.4)
Combination	173 (6.1)
Self-reported regular attender – n (%)	2504 (87.8)
Type of toothbrush – n (%)	
Manual	1363 (47.8)
Electric	1094 (38.3)
Both	171 (6.0)
Patient has dental insurance – n (%)	81 (2.8)
Smoked in the last 12 months – n (%)	320 (11.2)
How difficult is for you to travel to the dentist	6.4 (1.1), 2825
No. of scale and polishes in the year prior to the trial (routine data)	2.6 (2.3), 2688
Self-reported oral health	
I had bleeding from my gums when brushing – n (%)	453 (15.9)
Have you had bleeding from your gums when brushing your teeth?	2.1 (1.0), 2749
Have you had bad breath?	1.9 (0.9), 2576
Have you had an unpleasant taste in your mouth when brushing your teeth?	1.4 (0.7), 2584
Patient reported	
OHIP-14	5.2 (7.0), 2562
Oral health behaviour score	5.3 (1.7), 2636
After brushing I spit but not rinse	1060 (37.2)
Perceived behaviour control	4.6 (1.4), 2635
Attitude	4.8 (1.4), 2636
Clinical	
Calculus (% sites)	36.9 (27.4), 2849
Pocket depth (mm)	2.0 (0.4), 2843
Number of teeth	23.6 (4.7), 2853

gingivitis. We have developed the first diagnostic tool for gingival inflammation using a large UK-wide sample, combining self-reported oral health questions with recognized gingival inflammation risk factors. In this process, we have used the most up-to-date recommendations and reporting guidance (Collins et al., 2015; Cohen et al., 2016) and included expert discussion to identify potentially relevant risk factors.

Self-reported bleeding questions are preferable to questions focusing on gingival inflammation side-effects: they yielded better diagnostic performance and had lower rates of missing data, suggesting they are more acceptable to patients. These findings

are in line with previous studies from Gilbert and Nuttall (Gilbert & Nuttall, 1999) and Dietrich (Dietrich et al., 2005). Even though halitosis has been shown to be associated with bleeding gums (Kayombo & Mumghamba, 2017), it is possible that the indirect nature of the questions or their potential negative connotation is unhelpful.

A single self-reported bleeding question (Likert item) had comparable diagnostic accuracy performance to the diagnostic model we developed with four predictors. The diagnostic model had marginally better sensitivity and better specificity than the self-reported bleeding Likert item, resulting in a higher, moderate discriminant ability (Model's C-statistic = 0.65 vs Likert item area under the curve (AUC) = 0.60). Adding clinical predictors had a limited impact in the model's diagnostic performance. In practice, researchers that can access the risk factors information included in our model will benefit marginally from it compared with asking patients about their bleeding gums with a single Likert item, but this benefit is unlikely to be of clinical importance.

Smoking status, oral health behaviour, previous pattern of having scale and polishes and self-reported bleeding were selected as predictors in the final diagnostic model. The association between smoking and periodontal health is well established (Abbood et al., 2016; Du et al., 2018), and previous periodontitis prediction models had found oral health behaviour and dental visits as significant predictors (Du et al., 2018).

The original studies (IQuaD and INTERVAL) were not specifically designed as diagnostic accuracy or diagnostic model development studies, although we followed recommended practice where possible, including blinding assessors to index test results. Candidate predictors available were limited to those collected in IQuaD and INTERVAL and probably excluded important predictors that could have improved the diagnostic accuracy of the prediction model. SING's outcome, gingival inflammation, is challenging for two reasons: its definition and its measurement. We used gingivitis most recent threshold classification (Chapple et al., 2018) as an indicator of gingival inflammation. IQuaD and INTERVAL included intensive measurement training of outcome assessors and examined two tooth surfaces per tooth (partial mouth examination). This is recognized to be highly desirable for both patients and oral health professionals, and particularly in the context of two large UK-wide trials, even though the extent to which the results would differ from the results of a full-mouth examination is unknown (Trombelli et al., 2018). Bleeding on probing is impossible to calibrate. This may result in measurement error and misclassification of inflammation status which can lead to biased risk scores and c-statistics (Zawistowski et al., 2017). Because the extent of misclassification is unknown, we could not address it in our analysis, which is a common problem in this field (Kuchenhoff et al., 2009).

The diagnostic performance results found in SING are similar to other periodontitis prediction models (Du et al., 2018) and show challenges across the periodontal health field in identifying prediction models that are clinically useful. This could be due to a non-comprehensive

TABLE 3 Diagnostic accuracy estimates with 95% confidence intervals (CI) for each self-reported oral health question (reference standard: bleeding on probing $\geq 30\%$)

	Sensitivity (95% CI)	TP/N gingivitis ^a	Specificity (95% CI)	TN/N without gingivitis ^b	AUC (95% CI) ^c
Do your gums bleed when brushing?	0.20 (0.18, 0.22)	312/1564	0.88 (0.87, 0.90)	1088/1229	0.56 (0.54, 0.57)
Have you had bleeding from your gums when brushing your teeth?	0.73 (0.70, 0.75)	1111/1530	0.39 (0.36, 0.41)	471/1219	0.60 (0.58, 0.62)
Have you had bad breath?	0.64 (0.61, 0.66)	904/1420	0.40 (0.37, 0.42)	457/1156	0.52 (0.51, 0.54)
Have you had an unpleasant taste in your mouth when brushing your teeth?	0.31 (0.29, 0.34)	448/1423	0.72 (0.69, 0.75)	838/1161	0.52 (0.51, 0.54)

^aTrue positives divided by the number of people with clinical gingivitis.

^bTrue negatives divided by the number of people without clinical gingivitis.

^cArea under the curve (AUC) calculated using self-reported questions original Likert scale where applicable.

TABLE 4 Odds ratios and 95% confidence intervals (CI) of each predictor on gingivitis adjusted for all available predictors (final model)

Predictors	Odds ratio 95% CI, p-value
Bleeding gums (Likert scale)	1.40 (1.28,1.53), <.001
Non-smoker vs smoker (<i>reference category</i>)	1.73 (1.34,2.24), <.001
Scale and polish (previous year to the trial)	1.05 (1.01,1.08), .02
Oral health behaviour score	0.80 (0.76,0.84), <.001

identification of candidate predictors, but it is also related to heterogeneity in definition and measurement of disease and inflammation and absence of external validation of models.

There is space to improve the sensitivity of the diagnostic model developed here. Researchers have suggested other potential predictors to improve sensitivity of prediction models in this field: gingival cervical fluid, salivary markers or microbiology information have been identified as useful in periodontitis models (Du et al., 2018); systemic factors such as metabolic or nutritional factors and haematologic conditions can affect the extension, severity and progression of gingivitis and gingival inflammation (Chapple et al., 2018). However, it is important to find a balance between discrimination ability and cost of a diagnostic tool, as well as ease of implementation.

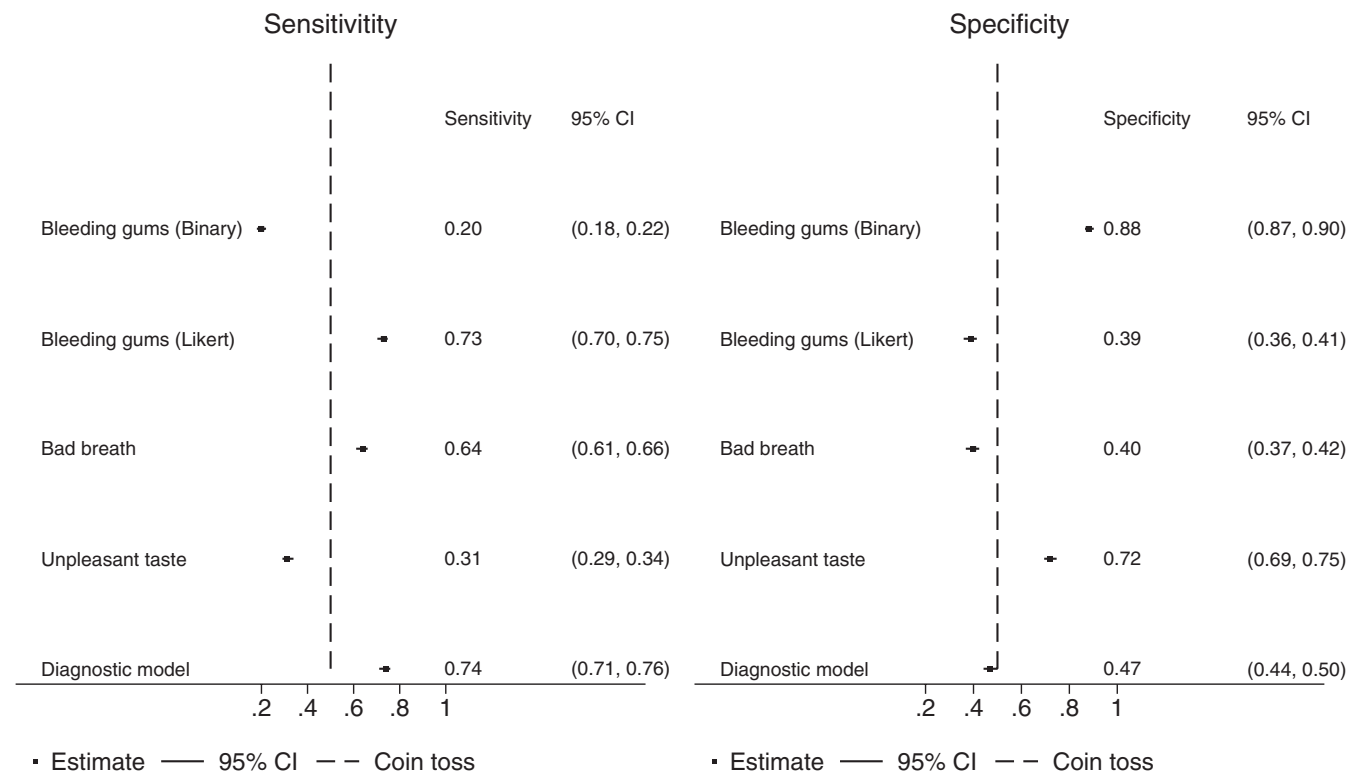


FIGURE 2 Sensitivity and specificity of the self-reported oral health questions and diagnostic model (CI, confidence interval)

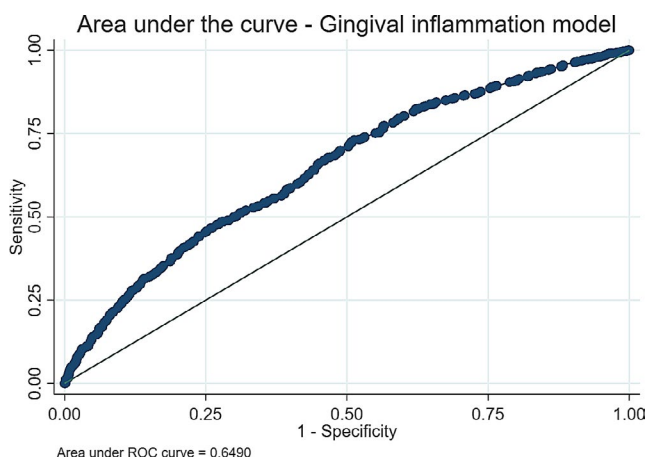


FIGURE 3 Area under the curve for the main model

The results reported in SING need to be replicated in different settings (when it comes to the diagnostic accuracy study) and externally validated (when it comes to the diagnostic model developed). SING included a large and representative sample of regular attenders to the dentist in the UK NHS, but it excluded participants with serious oral health disease and its results might not be generalizable to that population. On the one hand, participants attending the dentist regularly might be more aware of their oral health condition, but on the other hand participants suffering from more serious disease might be better at self-reporting it (Blicher et al., 2015).

We showed a simple self-reported bleeding gums question yielded similar but lower diagnostic accuracy to a diagnostic model including clinical and patient factors. Its sensitivity (0.73) is good and its specificity is low (0.39), so it needs to be improved to consider replacing clinical measures with self-reported ones. Including other risk factors improved diagnostic accuracy, but not in a clinically significant way. There are two possible ways forward: the improvement of self-reported bleeding gums questions by involving patients in their development and refinement and by collecting information closer to real-time, using health apps, which has recently showed promising results in identifying gingival inflammation using self-reported bleeding (Tonetti et al., 2020); the improvement of diagnostic accuracy in future diagnostic models to identify gingival inflammation by involving appropriate stakeholders in the identification of candidate predictors. As it currently is, the self-reported bleeding gums Likert item or the model is helpful to identify people with gingival inflammation and they should be considered if unfeasible or impossible to collect clinical data, for example in multiple imputation models if patients have missing clinical data.

ACKNOWLEDGEMENTS

We would like to acknowledge all study participants in IQuaD and INTERVAL for contributing with their information, and the trial teams involved in planning, conducting and delivering both trials.

CONFLICT OF INTEREST

The authors have stated explicitly that there are no conflicts of interest in connection with this article.

AUTHOR CONTRIBUTIONS

BG, GM and CR designed the study. CR co-led the data collection process. BG analysed the data. All authors were involved in the interpretation of results. BG drafted the manuscript. GM and CR commented on the manuscript. All authors agreed to the final version of the manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Beatriz Goulão  <https://orcid.org/0000-0003-1490-7183>

REFERENCES

- Abbood, H., Hinz, J., Cherukara, G., & Macfarlane, T. (2016). Validity of self-reported periodontal disease: A systematic review and meta-analysis. *Journal of Periodontology*, 87(12), 1474–1483.
- Blicher, B., Joshupura, K., & Eke, P. (2015). Validation of self-reported periodontal disease: A systematic review. *Journal of Dental Research*, 84(10), 881–890.
- Chapple, I. L. C., Dommisch, H., Glogauer, M., Johnson, G. K., Mealey, B. L., Eickholz, P., Geisinger, M. L., Genco, R. J., Glogauer, M., Goldstein, M., Griffin, T. J., Holmstrup, P., Johnson, G. K., Kapila, Y., Lang, N. P., Meyle, J., Murakami, S., Plemons, J., Romito, G. A., ... Yoshie, H. (2018). Periodontal health and gingival diseases and conditions on an intact and a reduced periodontium: Consensus report of workgroup 1 of the 2017 World Workshop on the Classification of Periodontology and Peri-Implant Diseases and Conditions. *Journal of Periodontology*, 89(Suppl 1), S74–84.
- Clarkson, J. E., Pitts, N. B., Bonetti, D., Boyers, D., Braid, H., Elford, R., Fee, P. A., Floate, R., Goulão, B., Humphris, G., Needleman, I., Norrie, J. D. T., Ord, F., van der Pol, M., Ramsay, C. R., Ricketts, D. N. J., Worthington, H. V., & Young, L., INTERVAL Trial Collaboration. (2018). INTERVAL (Investigation of NICE technologies for enabling risk-variable-adjusted-length) dental recalls trial : a multicentre randomised controlled trial investigating the best dental recall interval for optimum, cost-effective maintenance of oral health in dentate adults attending dental primary care. *BMC Oral Health*, 18(1), 135.
- Clarkson, J. E., Ramsay, C. R., Averley, P., Bonetti, D., Boyers, D., Campbell, L., Chadwick, G. R., Duncan, A., Elders, A., Gouick, J., Hall, A. F., Heasman, L., Heasman, P. A., Hodge, P. J., Jones, C., Laird, M., Lamont, T. J., Lovelock, L. A., Madden, I., ... Young, L. (2013). IQuaD dental trial; improving the quality of dentistry: a multicentre randomised controlled trial comparing oral hygiene advice and periodontal instrumentation for the prevention and management of periodontal disease in dentate adults attending dental primary care. *BMC Oral Health*, 13, 58.
- Cohen, J. F., Korevaar, D. A., Altman, D. G., Bruns, D. E., Gatsonis, C. A., Hooft, L., Irwig, L., Levine, D., Reitsma, J. B., de Vet, H. C. W., & Bossuyt, P. M. M. (2016). STARD 2015 guidelines for reporting diagnostic accuracy studies: Explanation and elaboration. *British Medical Journal Open*, 6(11), 1–17.
- Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for

- individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Annals of Internal Medicine*, 162, 55–63.
- Dietrich, T., Stosch, U., Dietrich, D., Schamberger, D., J-p, B., The, J. K., & Joshipura, K. (2005). The accuracy of individual self-reported items to determine periodontal disease history. *European Journal of Oral Sciences*, 113, 135–140.
- Du, M., Bo, T., Kapellas, K., & Peres, M. (2018). Prediction models for the incidence and progression of periodontitis: a systematic review. *Journal of Clinical Periodontology*, 45, 1408–1420.
- Eke, P., Thornton-Evans, G., Dye, B., & Genco, R. (2012). Advances in surveillance of periodontitis: the centres for disease control and prevention periodontal disease surveillance project. *Journal of Periodontology*, 83(11), 1337–1342.
- Gilbert, A. D., & Nuttall, N. M. (1999). Self-reporting of periodontal health status. *British Dental Journal*, 186(5), 241–244.
- Kallio, P. (1996). Self-assessed bleeding in monitoring gingival health among adolescents. *Community Dentistry and Oral Epidemiology*, 24, 128–133.
- Kayombo, C. M., & Mumghamba, E. G. (2017). Self-reported halitosis in relation to oral hygiene practices, oral health status, general health problems, and multifactorial characteristics among workers in ilala and temeke municipalities, Tanzania. *International Journal of Dentistry*, 2017, 1–10.
- Kuchenhoff, H. (2009). Misclassification and measurement error in oral health. In E. Lessafre, J. Feine, B. Leroux, & D. Declerck (Eds.), *Statistical and Methodological Aspects of Oral Health Research* (pp. 279–294). John Wiley & Sons, Ltd.
- Loe, H. (1967). The gingival index, the plaque index and the retention index systems. *Journal of Periodontology*, 38(6), 610–616.
- Newcombe, R. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in Medicine*, 17(8), 857–872.
- Petersen, P., & Ogawa, H. (2000). The global burden of periodontal disease: towards integration with chronic disease prevention and control. *Periodontology 2000*, 2012(60), 15–39.
- Ramos, R., Bastos, J., & Peres, M. (2013). Diagnostic validity of self-reported oral health outcomes in population surveys: literature review Validade diagnóstica de agravos. *Revista Brasileira de Epidemiologia = Brazilian Journal of Epidemiology*, 16(3), 716–728.
- Ramsay, C. R., Clarkson, J. E., Duncan, A., Lamont, T. J., Heasman, P. A., Boyers, D., Goulão, B., Bonetti, D., Bruce, R., Gouick, J., Heasman, L., Lovelock-Hempleman, L. A., Macpherson, L. E., McCracken, G. I., McDonald, A. M., McLaren-Neil, F., Mitchell, F. E., Dt Norrie, J., van der Pol, M., ... Young, L. (2018). Improving the Quality of Dentistry (IQuaD): a cluster factorial randomised controlled trial comparing the effectiveness and cost-benefit of oral hygiene advice and/or periodontal instrumentation with routine care for the prevention and management of periodontal disease in dentate adults attending dental primary care. *Health Technology Assessment: HTA / NHS R&D HTA Programme*, 22(38), 1–44.
- StataCorp. (2016). *Stata Statistical Software: Release 16*. : Statacorp LP.
- Tonetti, M. S., Deng, K., Christiansen, A., Bogetti, K., Nicora, C., Thurnay, S., & Cortellini, P. (2020). Self-reported bleeding on brushing as a predictor of bleeding on probing: Early observations from the deployment of an internet of things network of intelligent power-driven toothbrushes in a supportive periodontal care population. *Journal of Clinical Periodontology*, 47(10), 1219–1226.
- Trombelli, L., Farina, R., Silva, C. O., & Tatakis, D. N. (2018). Plaque-induced gingivitis: Case definition and diagnostic considerations. *Journal of Clinical Periodontology*, 45(Suppl 20), S44–67.
- Weintraub, J., Finlayson, T., Gansky, S., Santo, W., & Ramos-Gomez, F. (2013). Clinically determined and self-reported dental caries status during and after pregnancy among low-income hispanic women. *Journal of Public Health Dentistry*, 73(4), 311–320.
- Zawistowski, M., Sussman, J. B., Hofer, T. P., Bentley, D., Hayward, R. A., & Wiitala, W. L. (2017). Corrected ROC analysis for misclassified binary outcomes. *Statistics in Medicine*, 36(13), 2148–2160.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Goulão B, MacLennan GS, Ramsay CR. Have you had bleeding from your gums? Self-report to identify giNGival inflammation (The SING diagnostic accuracy and diagnostic model development study). *J Clin Periodontol*. 2021;48:919–928. <https://doi.org/10.1111/jcpe.13455>