Published in final edited form as:

J Clin Periodontol. 2024 May ; 51(5): 547–557. doi:10.1111/jcpe.13938.

Development and Validation of a Rule-based Algorithm to Identify Periodontal Diagnosis using Structured Electronic Health Record Data

Bunmi Tokede, DDS, MPH, DMSc [Associate Professor],
Department of Diagnostic and Biomedical Sciences, University of Texas at Houston, Health Science Center, Houston, Texas

Ryan Brandon, MSc [Consultant and Data Scientist],
Willamette Dental Group and Skourtes Institute, Hillsboro, OR, USA

Chun-Teh Lee, DDS, MS, DMSc [Associate Professor],
Department of Periodontics & Dental Hygiene, The University of Texas Health Science Center at Houston School of Dentistry, 7500 Cambridge Street, Suite 6470

Guo-Hao Lin, DDS, MS [Associate Professor],
Director, Postgraduate Periodontics Program, School of Dentistry, University of California, San Francisco, 707 Parnassus Avenue, D-3015, San Francisco, CA 94143

Joel White, DDS, MS [Professor],
Preventive and Restorative Dental Sciences, University of California, San Francisco/ UCSF School of Dentistry, San Francisco, CA, USA, 707 Parnassus Avenue, D-3248, Box 0758, San Francisco, CA 94143

Alfa Yansane, PhD [Associate Professor],
Preventive and Restorative Dental Sciences, University of California, San Francisco/ UCSF School of Dentistry, San Francisco, CA, USA, 3333 California Street, Ste. 495, San Francisco, CA 94118

Xiaoqian Jiang, PhD [Professor],
UTHealth School of Biomedical informatics, 7000 Fannin St Suite 600, Houston, Texas 77030

Elsbeth Kalenderian, DDS, MPH, PhD [Professor],
Department of Preventive and Restorative Dental Sciences, UCSF School of Dentistry, 707 Parnassus Ave., D-1025, Box 0758 | San Francisco, CA 94143

Muhammad Walji, PhD
Professor/Associate Dean for Technology Services and Informatics, Diagnostic and Biomedical Sciences Department, University of Texas Health Science Center at Houston, School of Dentistry, Houston, TX, USA, 7500 Cambridge, SOD 4184, Houston, TX 77054

First/corresponding author: Bunmi Tokede, DDS, MPH, DMSc, Associate Professor, Department of Diagnostic and Biomedical Sciences, University of Texas at Houston, Health Science Center, Houston, Texas, oluwabunmi.tokede@uth.tmc.edu.
Shared first author: Ryan Brandon
Shared senior authors: Elsbeth Kalenderian, Muhammad Walji

Authors declare no conflicts of interest

Abstract

Aim: To develop and validate an automated electronic health record (EHR)-based algorithm to suggest a periodontal diagnosis based on the 2017 World Workshop on the Classification of Periodontal Diseases and Conditions.

Methods:

Development: Using material published from the 2017 World Workshop, a tool was iteratively developed to suggest a periodontal diagnosis based on clinical data within the EHR. Pertinent clinical data included clinical attachment level (CAL), gingival margin to cementoenamel junction (GM-CEJ), probing depth (PD), furcation involvement (if present), and mobility.

Validation: Chart reviews were conducted to confirm the algorithm's ability to accurately extract clinical data from the EHR, and then to test its ability to suggest an accurate diagnosis. Subsequently, refinements were made to address limitations of the data, and specific clinical situations. Each refinement was evaluated through chart reviews by expert periodontists at the study sites.

Results: 323 charts were manually reviewed, and a periodontal diagnosis (healthy, gingivitis or periodontitis including stage and grade) was made by the expert periodontists for each case. After developing the initial version of the algorithm using the unmodified 2017 World Workshop criteria, accuracy was 71.8% for stage alone and 64.7% for stage and grade. Subsequently, 16 modifications to the algorithm were proposed, 14 were accepted. This refined version of the algorithm, had 79.6% accuracy for stage alone, and 68.8% for stage and grade together.

Conclusion: Our findings suggest that a rule-based algorithm for suggesting a periodontal diagnosis using EHR-recorded data can be implemented with moderate accuracy in support of chairside clinical diagnostic decision-making especially for inexperienced clinicians. Gray-zone cases still exist where clinical judgement continues to be required. Future applications of similar algorithms, with improved performance, is contingent on the quality (completeness/accuracy) of EHR data.

Keywords

Periodontitis; clinical decision support; diagnostic classification

INTRODUCTION

Periodontitis, a multifactorial inflammatory disease characterized by gingival inflammation and alveolar bone loss around teeth, has a high prevalence, affecting almost half of adults aged 30 or older in the United States and is the leading cause of tooth loss.[1]

It has long since been acknowledged that a classification scheme for periodontal and peri-implant diseases and conditions facilitates proper diagnosis and subsequent appropriate treatment, and is necessary for scientists to investigate etiology, pathogenesis, natural history, and treatment of those diseases and conditions.[2] Making an accurate periodontal diagnosis is important to promote timely interventions to prevent progression of disease. A new classification of periodontal and peri-implant diseases and conditions was introduced in

2018 based on the consensus reports of the 2017 international World Workshop organized by the American Academy of Periodontology (AAP) and the European Federation of Periodontology (EFP).[3] This classification was intended to redefine periodontitis by a multidimensional staging and grading system.[4] Staging is largely dependent upon the severity of disease at presentation as well as on the complexity of disease management, while grading provides supplemental information about biological features of the disease, including a history-based analysis of the rate of disease progression, assessment of the risk for further progression, anticipated poor outcomes of treatment, and assessment of the risk that the disease or its treatment may negatively affect the general health of the patient.[5]

As can be implied from the foregoing, staging is determined after considering mostly objective and readily available cross-sectional variables. Grading, conversely, is determined by information that rely on longitudinal assessments, for which data availability becomes an issue.

Compared to the previous classification introduced in 1999,[6] the AAP/EFP 2017 World Workshop classification of periodontitis considers more clinical parameters, such as radiographic bone loss (RBL), probing depth (PD) and furcation involvement, relevant to clinical practice. However, these additional clinical parameters also make the classification more sophisticated which is challenging for less experienced dentists, dental hygienists and dental students to learn.[7–9] Therefore, a periodontal diagnosis clinical decision support (CDS) tool integrated within an electronic health record (EHR) can potentially assist care providers in arriving at an accurate periodontal diagnosis.

This study aimed to develop and validate an algorithm using structured data from the EHR to suggest a periodontal diagnosis based on periodontal charting, clinical findings, dental and medical history, using the 2017 World Workshop classification of periodontal diseases.

METHODS

The study was conducted in accordance with the guidelines of the World Medical Association's Declaration of Helsinki, and approved by the University of California, San Francisco School of Dentistry (UCSF) and the University of Texas Health Science Center at Houston (UTHealth) Committee for the Protection of Human Subjects (HSC-DB-21-0616).

EHR-based Algorithm Development

The development of the algorithm was carried out by a multi-institutional, multi-disciplinary team. Using materials and documentation from the AAP and EFP, an initial flowchart of the algorithm was adapted from previous work (Figure 1). [10] A Structured Query Language (SQL) script was developed to extract the clinical data elements to be used in the algorithm. These elements included Bone Loss (BL), Clinical Attachment Loss (CAL), Pocket Depth (PD), Bleeding on Probing (BOP), tooth loss due to periodontitis, number of present teeth/ occluding pairs, furcation involvement and tooth mobility.

Additional SQL procedures were developed to follow the logic of the flowchart to compute a periodontal diagnosis based on the clinical data available for a specific patient and date. The algorithm suggested a diagnosis of Health, Health on reduced periodontium, Gingivitis, Gingivitis on reduced periodontium or Periodontitis. In presenting our results and analyses, the first four diagnoses were combined as 'not periodontitis'. For cases classified as periodontitis, the automated algorithm followed the 2017 World Workshop AAP/EFP guidelines to classify each case as Stage I, II, III or IV, and to assign the appropriate disease grade of A, B or C.

Stages I through IV are generally assigned according to the degree of clinical attachment loss, amount of radiographic bone loss, and tooth loss due to periodontitis. Probing depth and other complexity factors including presence of vertical bone loss ≥3mm, furcation involvement class II/III, secondary occlusal trauma (tooth mobility degree≥ II), and number of remaining teeth (opposing pairs) were also considered in the assessment of a Stage III or IV case. Ridge deficiency, masticatory dysfunction, bite collapse, drifting, and flaring were not included in the SQL logic, given these factors are not routinely well documented in the EHR. Grading was assigned as described by the 2017 World Workshop [5], assuming a moderate rate of progression (grade B) and then looking for direct and indirect measures of the progression in the past as a means of improving the establishment of prognosis for the individual patient. These considered measures of progression included changes in radiographic bone level or CAL, smoking status, diabetes and the assessment of radiographic bone loss in relation to patient age.

Diagnoses of reduced periodontium were determined following the consensus report of periodontal health and gingival diseases and conditions on an intact and reduced periodontium.[11]

Iterative Refinement—Using retrospective patient-level data, we implemented the clinical data extract and periodontal diagnosis algorithm to flag cases for review at the two participating institutions (UTHealth and UCSF). Both institutions use the same EHR platform (axiUm, Exan, Vancouver, Canada) and the Systematized Nomenclature for Dental Diagnostic System (SNODDS) diagnostic terminology integrated within the EHR.[12]

Subsequently, enhancements and refinements to the algorithm were made to address limitations of the data, to address complexity factors of secondary occlusal trauma and occluding pairs, and to confirm the presence of CAL related specifically to periodontal disease. Additional refinements were implemented to confirm some undocumented but underlying assumptions of the publications. These included the exclusion of third molars and implant sites from the assessment of teeth with PD or CAL and the requirement of at least two affected teeth.[13] The itemized adjustments made to improve the accuracy and reliability of the algorithm are shown in the results section below.

Following each round of revisions to the algorithm, more patient charts were reviewed to see if/how the revision altered the suggested diagnosis. Then, a consensus process between the reviewers determined if the revision improved the accuracy of the algorithm and should be adopted. With this process, there were two proposed revisions that were rejected and not implemented in the finalized algorithm. This iterative process resulted in the algorithm described in Figure 2.

Chart Reviewers—Two primary reviewers (C.L, G.L) independently reviewed the included charts. Both are clinician-scientists with an in-depth knowledge in the etiology and treatment of periodontal disease. They are Board-certified by the American Board of Periodontology and have extensive experience (> 15y) in providing complex patient care and conducting clinical research as full-time clinical faculty. Each serves as director of postgraduate periodontology at their respective institutions.

Algorithm Validation

After the algorithm was finalized, we performed chart reviews on a representative sample of patient charts at each of the institutions. Data were mined from the EHR of each site for the period January 1, 2021, through December 31, 2021. Charts included were limited to cases from resident periodontology programs, with requirements that each case had complete periodontal charting including PD, CAL and the distance from the gingival margin to the cemento-enamel junction (GM-CEJ). Cases were further limited to adult patients over the age of 16 with a minimum of 10 natural teeth present and having recent bitewing radiographs (within 6 months).

In performing these reviews, the algorithm-suggested diagnosis was compared to the diagnosis that was originally selected by the clinician at the point of care. Stage and grade of periodontitis diagnosis were considered. Cases in which the suggested diagnosis and the clinician diagnosis were identical were referred to as matched cases. Cases in which these diagnoses were at variance were designated as non-matched cases. This validation phase of the algorithm included the manual review of both matched and non-matched cases. Challenging cases that the primary reviewers could not decide the diagnosis were discussed in a panel made up of senior clinicians and specialists. Two national leaders in the field of periodontology were consulted in a few cases in which panelists could not agree on a diagnostic classification.

Primary reviewers evaluated a randomly selected sample of matching and non-matching cases (oversampling for the non-matched cases – two-thirds non-matching and one-third matching) to assess the performance of the algorithm across all variations of periodontal disease. Reviewers had access to radiographs to establish a gold-standard periodontal diagnosis for each case. To assess inter-rater reliability, a subset of reviewed charts at each site was independently reviewed by another periodontist.

Statistical Methods

Between the two sites there were 811 charts (UCSF: 424, UTH: 387) that satisfied the chart inclusion criteria.

In order to determine the minimum number of charts to be manually reviewed, the sample size calculation for categorical data (proportions) was utilized.[14] We set initial values for sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) proportions at 80%. Given a population sample of 811 EHR records, we set the margins for error and significance levels at each site at a standard two-tailed z-value 0.05 ($d = 0.05$; $z = 1.96$) and applied "Cochran's correction".[15] This yielded a sample of 189 needed to review. Additionally, the number of charts needed for manual review to establish inter-rater reliability between the reviewers was calculated. For 2 reviewers, a standard significance level of 0.05 and (1-beta) = 0.80, a difference between Kappa values of .2, the minimum number required to review was 39.[16]

Descriptive summary statistics were used to describe the patient demographics in terms of means and standard deviation for numeric variables, and frequency distributions and percentage contributions for categorical variables. We performed all statistical analyses by using software (R for Statistical Computing).

We characterized algorithm performance (validity) in terms of sensitivity, specificity, NPV and PPV, respectively, as well as the degree of agreement (kappa) between the two independent primary reviewers.

RESULTS

Algorithm Development (adopted refinements)

1. All third molars were excluded from diagnosis considerations due to the various conditions of third molars.
2. All implant sites were excluded from diagnosis considerations. PD or CAL on an implant site may imply specific implant-related disease but should not impact the patient's overall periodontal diagnosis.
3. In the absence of documented bone loss, furcation involvement of class II and/or III as recorded in the periodontal charting or structured clinical forms, was used as a proxy for bone loss to assess a potential periodontitis case.
4. In the absence of radiographic bone loss, at least two non-adjacent teeth with interdental CAL> 2mm was used to assess a potential periodontitis case. The CAL threshold was set to be >2mm instead of ≥1mm because overestimation of CAL was observed to be prevalent in periodontal chartings of the present data set. As such, the increased threshold reduced the number of cases where incorrectly charted CAL or CAL not related to periodontal disease was misclassifying the case as periodontitis.
5. CAL was used in the algorithm only if GM-CEJ was also recorded in the periodontal chart on the same site. CAL recorded without GM-CEJ measurement was excluded from this work.
6. Interdental CAL adjacent to an edentulous area was reduced by 2mm for the assessment (for example, 5mm value was reduced to 3mm for algorithm) to account for loss of bone due to the neighboring extraction site. Papillary recession and bone loss of the tooth adjacent to an extraction site is a known fact. Since the algorithm assesses periodontal conditions based on the results of the periodontal charting, the loss of attachment (higher CAL value) in these sites adjacent to the edentulous area may make the algorithm over-diagnose periodontal diseases. [17, 18]

7. CAL on the distal site of a second molar was reduced by 2mm unless the adjacent third molar was fully erupted. This supported accommodation for CAL due to a neighboring partially erupted third molar or history of third molar extraction.
8. Sites with 4mm PD and no BOP were considered healthy Sites with PD>4mm or 4mm with BOP were included as contributing to the diagnosis of periodontitis. The history of periodontal treatment is not considered. This change addressed patients presenting with 4mm PD and no BOP that have previously been successfully treated for periodontitis.
9. To determine a periodontitis case of Stage III/IV, interdental CAL≥5mm in more than 2adjacent teeth was modified to remove the adjacency requirement and change the required number of teeth involved. The revised condition of CAL≥5mm in at least 2 or more teeth (adjacent or non-adjacent) was used. This adjustment can include some severe periodontitis cases with localized disease.
10. To consider a periodontitis case as Stage III/IV, the presence of PD≥6mm in more than 2 adjacent teeth was modified to remove the adjacency requirement and change the required number of teeth involved. The revised condition of PD≥6mm in at least 2 or more teeth was used. This adjustment can include some severe periodontitis cases with localized disease.
11. When assessing the Stage IV complexity requirement of 10 occluding pairs, both natural teeth and occluded crowns (including implant supported crowns and pontics) were considered in occlusion with the opposing arch/tooth. This refinement was made because the original criterion is related to the need for rehabilitation.
12. For assessment of the Stage IV complexity requirement of secondary occlusal trauma, if 5 or more teeth are present with mobility degree≥2 and the case has fulfilled the criteria of Stage III, then it was upgraded to Stage IV. This refinement was made because secondary occlusal trauma was not well documented in our records.
13. To consider a case for Periodontitis, a requirement of two or more teeth with PD (PD 4mm with BOP or PD > 4mm) was implemented. This reduced the number of cases classified as Periodontitis due to just a single tooth with PD.

Algorithm Development (rejected refinements)

1. Reducing the threshold for interdental CAL to ≥ 2mm for assessment of a Periodontitis diagnosis. After review of the cases with a maximum CAL of 2mm, it was determined that the majority of these cases represented CAL that was either recorded incorrectly in the EHR or CAL that was not due to periodontal disease. The threshold was set to ≥3mm as noted in the figure 2.
2. Adjacent teeth vs. Adjacent probing sites: for the assessment of 2 or more non-adjacent *sites* with CAL ≥ 3mm. After reviewing cases, the consensus was to use adjacent teeth for this assessment.

Algorithm Validation/Performance

For this, a total of 323 charts were reviewed at the 2 sites (UCSF: 122, UTH: 201). 214 of the reviews (66.3%) were performed on complex "non-matching" cases where the algorithm suggested diagnosis did not agree with the point-of-care assessment of the case. Mean patient age was 54.4 (SD = 15.2), 57.9% identified as female. The most frequently identified racial category was "White" (26.3%), though many did not self-report their race (Table 1). A plurality of sample patients paid for treatment in cash (62.6%), followed by government insurance/Medicaid (22.9%), private insurance (14.5%). 87.4% identified as non-smokers and 15.0% had a documented diagnosis of diabetes. Table 2 shows the number of included patients by periodontal diagnostic stage/extent as assessed by the primary reviewers.

10% of the originally reviewed charts were re-reviewed by each provider to evaluate intra-rater reliability. Percent agreement for disease stage was 100% for both reviewers.

Tables 3 and 4 show the performance of algorithm-generated diagnosis (stage only or stage + grade) against the periodontist reviewed diagnosis. Diagnoses of healthy periodontium, healthy on reduced periodontium, and gingivitis were all combined and designated as "not periodontitis". Both tables show the comparative results for disease stage between the de-novo algorithm (using the AAP diagnostic criteria – figure 1) and the refined algorithm (shown in figure 2).

Table 3 shows the 'stage' concordance between the algorithm and the periodontists for the de-novo was 71.8% (95% CI: 66.6, 76.7). The Kappa correlation coefficient was .59 which represents "moderate" agreement. The sensitivity, specificity, positive predictive value, and negative predictive value are reported by stages (I, II, III, IV). The Kappa correlation for the modified algorithm was 0.70, representing substantial agreement between the modified algorithm and the gold standard.

Table 4 shows the 'stage + grade' concordance between the algorithm and the periodontists for the de-novo was 64.7% (95% CI: 59.1, 69.9). The Kappa correlation coefficient was 0.53 which represents "moderate" agreement. The sensitivity, specificity, positive predictive value, and negative predictive value are reported by stages (I, II, III, IV). The Kappa correlation for the modified algorithm was 0.58, also representing moderate agreement between the modified algorithm and the gold standard.

The difference in the performance of the de novo vs refined algorithm was significant for staging but not significant when the stage and grade were combined (Table 5).

An additional 46 charts were reviewed to test for inter-rater reliability with each reviewer reviewing a subset of the charts reviewed by the other primary reviewer. The percent agreement was 78.3%. The inter-rater reliability was 0.69 (weighted $\kappa = 0.69$, 95% CI: 0.51, 0.87). This Kappa coefficient was statistically significantly different from zero ($z = 7.56$, p-value = <0.0001).

DISCUSSION

The current results demonstrate moderate accuracy of an EHR-based algorithm in suggesting a clinical periodontal diagnosis. This algorithm can be integrated within a clinical decision support tool in an EHR system to assist clinicians in making a periodontal diagnosis at the point of care.[10] Or to identify potential diagnostic errors; cases where documented diagnoses may be inconsistent with the underlying clinical data. Equally important is the finding that "gray zones" exist in real world clinical practice in which clinical judgement will be more appropriate than using just clinical data to determine the periodontal diagnosis.[13]

When looking at disease stage only, the refined algorithm performed significantly better than the de novo algorithm (Tables 2 and 4). However, when stage and grade of disease were combined, the modified algorithm only performed marginally better, and this difference was not statistically significant (Tables 3 and 4). This is probably due to the fact that staging is determined after considering mostly readily available cross-sectional variables. Grading, conversely, is determined by information that rely on longitudinal assessments, for which data availability becomes an issue.

To improve the tool's accuracy in suggesting a diagnosis across all levels of disease, our team primarily focused reviews on the more complicated cases where the tool did not agree with the diagnosis recorded at the point of care. The distribution of included cases (by stage) and level of inter-rater reliability reflects the complexity of the test cases reviewed in the study and is comparable to that reported in a previous study.[19] The original classification system performed well on simple cases, but in comprehensive review of many "grey-zone" cases, we were able to identify and address limitations of the EHR data that led to a more accurate tool across both simple and complex cases.

Due to the high prevalence of minor attachment loss (CAL=1 or 2mm) documented in patients without actual RBL, the CAL threshold for determining a periodontitis case was modified in the CDS tool. It is also known that the measurement accuracy of CAL for inexperienced clinicians might be questionable.[20] At the participating dental schools, most of the periodontal chartings were completed by periodontal residents with limited years of experience in practice. Therefore, the distance of GM-CEJ was often underestimated for healthy sites when the CEJ was located sub-gingivally, and consequent overestimation of the attachment loss. Due to this discrepancy, the decision was made to implement "interdental CAL>2mm" rule instead of interdental CAL >0mm to determine true periodontitis cases. This adjustment reduced the number of healthy or gingivitis cases from being incorrectly classified as periodontitis.

Furthermore, the authors decided to reduce the interdental CAL of the site adjacent to an edentulous area by 2mm in the tool logic, since attachment loss is a reasonable expectation adjacent to an extraction socket after tooth extraction. Similarly, the CAL at the distal site of a second molar was also reduced by 2mm because it is common that attachment loss at this specific site is associated with the presence of an impacted third molar or a history of third molar extraction.[21, 22]

It is worth noting that in the current study, the algorithm identified a very limited number of Stage I periodontitis cases (Table 1). As the algorithm had been modified to assign a periodontitis diagnosis only to sites with interdental CAL>2mm, and the CAL next to an edentulous area was also reduced by 2mm, it is not surprising that cases with only 1-2mm of CAL, which would have qualified for Stage I periodontitis diagnosis, were assigned a ‘not-periodontitis’ diagnosis. Even after manual review, only five Stage I periodontitis cases were identified. This could be because the charts included in the current study were from the periodontal residency clinics at the two centers. It is conceivable that most Stage I periodontitis patients were not represented in this sample because they are hardly referred to the residency clinics.

To determine a stage, at least two teeth (adjacent or non-adjacent) have to fulfill the criteria of the specific stage. This approach is to avoid overestimating the severity of a periodontitis case although observed periodontal destruction of a single tooth can determine the stage based on clinical judgement.[23]

Increased adoption of EHRs provides us the opportunity to efficiently extract useful data for measuring performance e.g. accuracy of periodontal diagnoses, assess the relationships between performance and health outcomes, and benchmark population health. EHRs already provide some access to public health data to study the population for potential health improvements and act as a safety net for potential health threats. In previous work[24] involving approximately half a million data points and 200,000 patients from the EHRs of 3 institutions, the authors were able to provide insight into what has been referred to as two of the most meaningful clinical end points in periodontology—stability of clinical attachment level (no new periodontitis diagnosis) and tooth survival (no new tooth loss). Hitherto, the clinical periodontology literature has mostly represented studies that use surrogate end points.[25] This highlights the potential of EHR data for research in periodontology and other domains in clinical dentistry.

Looking forward, AI in healthcare is still largely dominated by the development of expert systems (like the tool presented here), based on rules derived from experts, then translated and programmed. Machine learning (ML) algorithms are now being developed to overcome the constraints of expert systems.[26] In ML, engineers program algorithms able to derive their own rules from data. Thus, human-coded rules are replaced by machine-generated, data-driven rules. This allows ML systems to learn from data and interpret unknown situations. Among the panel of ML techniques developed, deep learning based on artificial neural networks is the most ubiquitous.[27] Diagnostic clinical decision making in periodontology will benefit from these advances in ML. Machine learning algorithms can be used to help identify patterns in data that may not be immediately apparent to the naked eye, and this can be extremely helpful in diagnosing and treating periodontal disease, which will benefit both patients and practitioners by providing more accurate and efficient diagnosis and treatment of periodontal disease.[28, 29]

LIMITATIONS

This study has limitations that should affect how the findings are interpreted. The periodontal diagnoses assessed by the expert reviewers were decided based on periodontal charting in addition to clinical report, intraoral radiographs and clinical judgement; but the tool-assigned diagnosis was determined only by the charting. Therefore, it is understandable that the accuracy of the algorithm generated diagnosis was imperfect in some cases. This is particularly true for Stage IV cases. Stage IV may be differentiated from Stage III by history of missing teeth attributable to periodontitis plus additional hopeless teeth. A Stage IV case should also present with features identified in the classification that define the need for complex rehabilitation, including masticatory dysfunction, bite collapse, and potential for losing major dentition components. These features are not well documented in our records. The algorithm was also validated using data from dental schools. As such, for generalizability, it has to be further evaluated using data from different clinical settings, such as private offices and group practices. A further note on generalizability – for the study year, only 42% (site 1) and 25% (site 2) of patients seen for a comprehensive periodontal evaluation had complete documentation of clinical periodontal charting in the specific cohort. Even fewer of those had complete medical information and radiological bone level documentation – all of which are critical in diagnosing and treating periodontal disease. Quality and quantity of documentation can limit the utility of having and deploying these algorithms and decision support tools.

CONCLUSION

Our findings suggest that methods for suggesting a periodontal diagnosis using EHR-recorded clinical data can be implemented with moderate accuracy. The reliability of the automated system is highly contingent upon the quality of the EHR data. Gray-zone cases exist where clinical judgement is required to determine a diagnosis. When accurate and complete clinical data is available in the EHR, an algorithm following the guidelines set forth by the World Workshop can be used to reduce clinician burden at the point of care, promote understanding of this nascent 2017 periodontal diagnostic classification system, and support advancement in quality assessment, patient care and research for a learning health system.

ACKNOWLEDGEMENT

Funding for this research was provided by the U.S. Department of Health and Human Services, Agency for Healthcare Research and Quality. Research Grant no.: R01HS027938 entitled: “Dental Diagnostic Failures”.

BT, RB, EK and MW conceived the study, were responsible for its execution, and wrote the first draft of the manuscript. CTL, GHL and JW helped with the iterative development and refinement of the algorithm. They also performed all chart reviews included in this study. AY and XJ conducted the statistical analyses and provided their expertise with the methodological approach for this work. All authors critically reviewed the manuscript.

We would also like to thank Dr. Gary Armitage for his help in clarifying some of the 2017 World Workshop diagnostic classification guidelines, and his help in arriving at a diagnosis in gray-zone cases.

REFERENCES

1. Eke PI, Thornton-Evans GO, Wei L, et al. Periodontitis in US Adults: National Health and Nutrition Examination Survey 2009-2014. J Am Dent Assoc 2018;149(7):576–88 e6. [PubMed: 29957185]
2. Kalenderian E, Ramoni RL, White JM, et al. The development of a dental diagnostic terminology. J Dent Educ 2011;75(1):68–76. [PubMed: 21205730]

3. Papapanou PN, Sanz M, Buduneli N, et al. Periodontitis: Consensus report of workgroup 2 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. J Periodontol 2018;89 Suppl 1:S173–S82. [PubMed: 29926951]
4. Tonetti MS, Greenwell H, Kornman KS. Staging and grading of periodontitis: Framework and proposal of a new classification and case definition. J Periodontol 2018;89 Suppl 1:S159–S72. [PubMed: 29926952]
5. Papapanou PN, Sanz M, Buduneli N, et al. Periodontitis: Consensus report of workgroup 2 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. Journal of periodontology 2018;89:S173–S82. [PubMed: 29926951]
6. Armitage GC. Development of a classification system for periodontal diseases and conditions. Ann Periodontol 1999;4(1):1–6. [PubMed: 10863370]
7. Oh SL, Yang JS, Kim YJ. Discrepancies in periodontitis classification among dental practitioners with different educational backgrounds. BMC Oral Health 2021;21(1):39. [PubMed: 33482794]
8. Abou-Arraj RV, Kaur M, Alkhoury S, et al. The new periodontal disease classification: Level of agreement on diagnoses and treatment planning at various dental education levels. J Dent Educ 2021;85(10):1627–39. [PubMed: 33955000]
9. Nibali L, Shemie M, Li G, et al. Periodontal furcation lesions: A survey of diagnosis and management by general dental practitioners. Journal of Clinical Periodontology 2021;48(11):1441–48. [PubMed: 34472119]
10. Tonetti MS, Sanz M. Implementation of the new classification of periodontal diseases: Decision-making algorithms for clinical practice and education. Journal of Clinical Periodontology 2019;46(4):398–405. [PubMed: 30883878]
11. Chapple IL, Mealey BL, Van Dyke TE, et al. 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 Periodontal and Peri-Implant Diseases and Conditions. Journal of periodontology 2018;89:S74–S84. [PubMed: 29926944]
12. Ramoni RB, Etolue J, Tokede O, et al. Adoption of dental innovations: The case of a standardized dental diagnostic terminology. The Journal of the American Dental Association 2017;148(5):319–27. [PubMed: 28364948]
13. Kornman KS, Papapanou PN. Clinical application of the new classification of periodontal diseases: Ground rules, clarifications and “gray zones”. J Periodontol 2020;91(3):352–60. [PubMed: 31891194]
14. Kotrlik J, Higgins C. Organizational research: Determining appropriate sample size in survey research appropriate sample size in survey research. Information technology, learning, and performance journal 2001;19(1):43.
15. Robb R. w. G. Cochran, Sampling Techniques (John Wiley & Sons, 1963), ix+ 413 pp., 72s. Proceedings of the Edinburgh Mathematical Society 1963;13(4):342–43.
16. Shoukri MM, Asyali M, Donner A. Sample size requirements for the design of reliability study: review and new results. Statistical methods in medical research 2004;13(4):251–71.
17. Chappuis V, Araújo MG, Buser D. Clinical relevance of dimensional bone and soft tissue alterations post-extraction in esthetic sites. Periodontology 2000 2017;73(1):73–83. [PubMed: 28000281]
18. Barone A, Ricci M, Tonelli P, Santini S, Covani U. Tissue changes of extraction sockets in humans: a comparison of spontaneous healing vs. ridge preservation with secondary soft tissue healing. Clinical oral implants research 2013;24(11):1231–37. [PubMed: 22784417]
19. Ravidà A, Travan S, Saleh MH, et al. Agreement among international periodontal experts using the 2017 World Workshop classification of periodontitis. Journal of periodontology 2021;92(12):1675–86. [PubMed: 34545953]
20. Vandana K, Gupta I. The location of cemento enamel junction for CAL measurement: A clinical crisis. Journal of Indian Society of Periodontology 2009;13(1):12. [PubMed: 20376234]
21. Chen Y-W, Lee C-T, Hum L, Chuang S-K. Effect of flap design on periodontal healing after impacted third molar extraction: a systematic review and meta-analysis. International Journal of Oral and Maxillofacial Surgery 2017;46(3):363–72. [PubMed: 27600798]

22. Lee C-T, Hum L, Chen Y-W. The effect of regenerative periodontal therapy in preventing periodontal defects after the extraction of third molars: A systematic review and meta-analysis. The Journal of the American Dental Association 2016;147(9):709–19. e4. [PubMed: 27085787]
23. Kornman KS, Papapanou PN. Clinical application of the new classification of periodontal diseases: Ground rules, clarifications and "gray zones". Journal of periodontology 2020;91(3):352–60. [PubMed: 31891194]
24. Tokede B, Yansane A, White J, et al. Translating periodontal data to knowledge in a learning health system. The Journal of the American Dental Association 2022;153(10):996–1004. [PubMed: 35970673]
25. Loos BG, Needleman I. Endpoints of active periodontal therapy. Journal of Clinical Periodontology 2020;47:61–71. [PubMed: 31912527]
26. Mitchell TM, Mitchell TM. Machine learning: McGraw-hill New York; 1997.
27. Hinton G Deep learning—a technology with the potential to transform health care. Jama 2018;320(11):1101–02. [PubMed: 30178065]
28. Lee CT, Kabir T, Nelson J, et al. Use of the deep learning approach to measure alveolar bone level. Journal of clinical periodontology 2022;49(3):260–69. [PubMed: 34879437]
29. Kearney VP, Yansane A- IM, Brandon RG, et al. A generative adversarial inpainting network to enhance prediction of periodontal clinical attachment level. Journal of Dentistry 2022;123:104211. [PubMed: 35760207]

Clinical Relevance Statement

Scientific Rationale:

To develop an electronic health record (EHR) rule-based algorithm that can help to reliably arrive at an accurate periodontal diagnostic classification (stage and grade), using the 2017 World Workshop diagnostic classification system.

Principal Findings:

Compared to expert assessment, the final algorithm had 69% accuracy. The modest performance was predominantly due to data limitations, complex cases and clinical grey areas.

Practical Implications:

An algorithm cannot yet replace expert diagnostic assessment for periodontal diagnostic classification. Gray-zone cases exist where clinical judgment is required. Future applications of similar algorithms remain contingent on the quality (completeness/ accuracy) of EHR data.

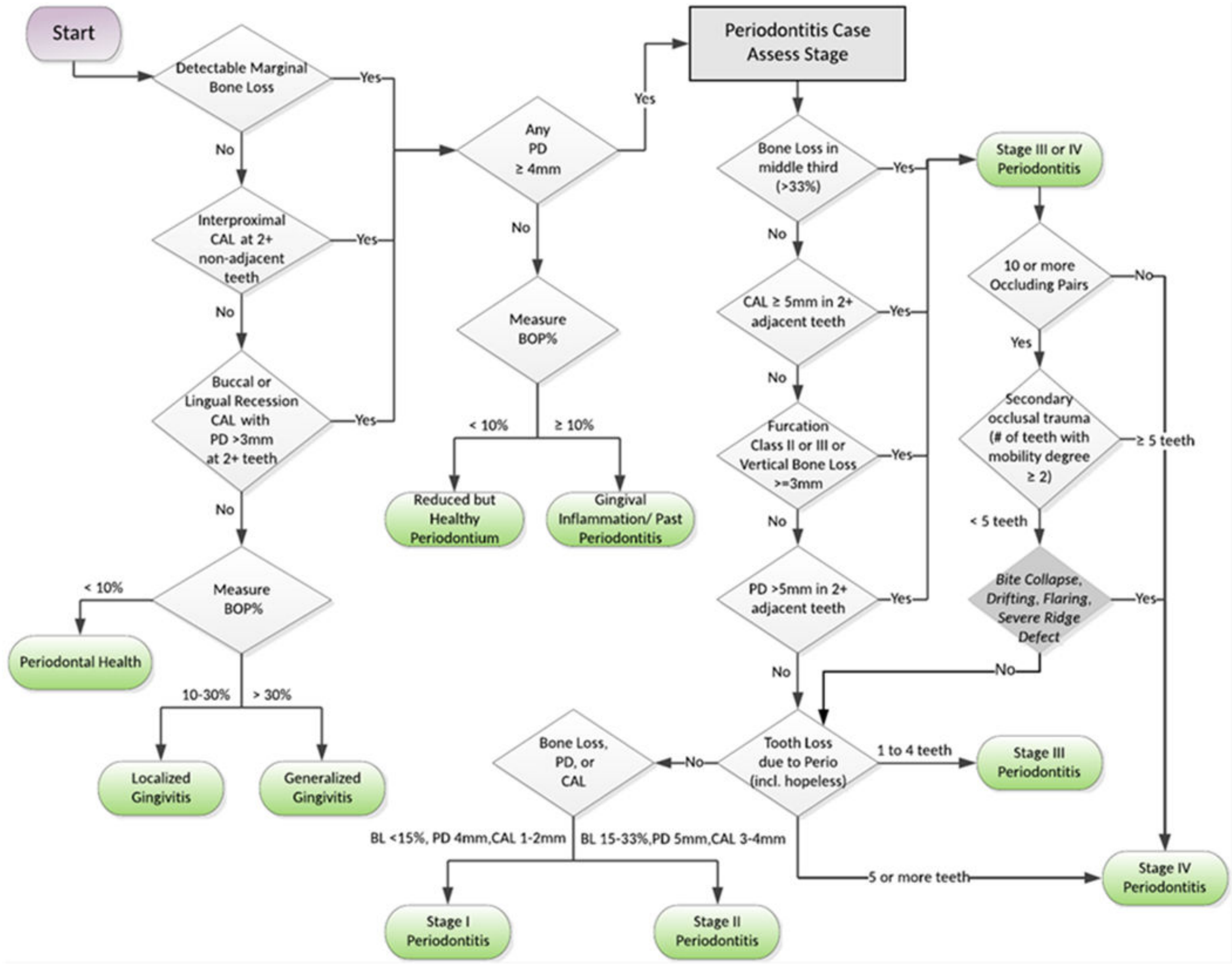

Figure 1:
Periodontal diagnosis algorithm flowchart for initial decision support tool [5, 10]
Bone Loss (BL), Clinical Attachment Loss (CAL), Pocket Depth (PD), Bleeding on Probing (BOP)

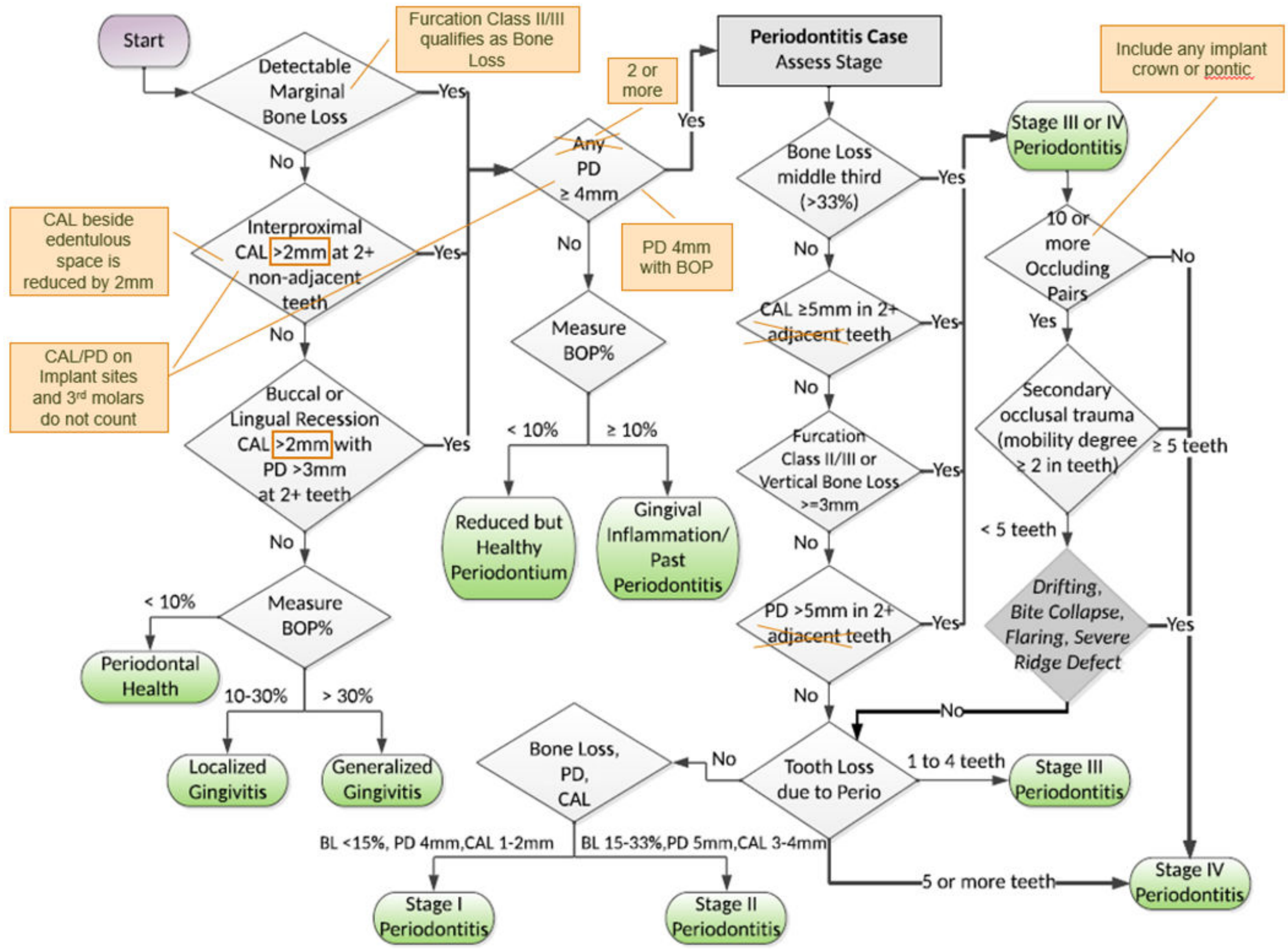

Figure 2:

Decision algorithm flowchart with approved edits (shown in yellow boxes).

CAL: Clinical Attachment Loss, PD: Pocket Depth, BOP: Bleeding on Probing

Table 1:

Patient Descriptive Statistics (N=323)

	SITE 1		**SITE 2**		**TOTAL**	
Count	201		122		323	
Age	55.5	14.4	52.5	15.4	54.3	14.8
Gender						
Male	79	39.3%	54	44.3%	133	41.2%
Female	121	60.2%	68	55.7%	189	58.5%
Other	1	0.5%	--	0.0%	1	0.3%
Race						
American Indian/Alaska Native	2	1.0%	--	0.0%	2	0.6%
Asian	13	6.5%	20	16.4%	33	10.2%
Black/African American	18	9.0%	8	6.6%	26	8.0%
Hawaiian/Pacific Islander	--	0.0%	--	0.0%	--	0.0%
White	54	26.9%	31	25.4%	85	26.3%
More than one race	7	3.5%	4	3.3%	11	3.4%
Unknown/Not reported	53	26.4%	25	20.5%	78	24.1%
Other	54	26.9%	34	27.9%	88	27.2%
Ethnicity						
Hispanic	8	4.0%	5	4.1%	13	4.0%
Non-Hispanic	3	1.5%	17	13.9%	20	6.2%
Unknown	190	94.5%	100	82.0%	290	89.8%
Insurance Type						
Cash	190	94.5%	28	23.0%	218	67.5%
Private	10	5.0%	13	10.7%	23	7.1%
Delta	--	0.0%	16	13.1%	16	5.0%
Denti-Cal	--	0.0%	65	53.3%	65	20.1%
Medicaid	1	0.5%	--	0.0%	1	0.3%
Smoking Status						
No	181	90.0%	104	85.2%	285	88.2%
Yes	12	6.0%	--	0.0%	12	3.7%
Yes/Heavy	3	1.5%	--	0.0%	3	0.9%
Yes/Light	5	2.5%	--	0.0%	5	1.5%
Unknown	0	0.0%	15	12.3%	15	4.6%
Diabetes Status						
No	178	88.6%	95	77.9%	273	84.5%

	SITE 1		SITE 2		TOTAL	
Yes	15	7.5%	9	7.4%	24	7.4%
Yes/Control	3	1.5%	--	0.0%	3	0.9%
Yes/Uncontrolled	5	2.5%	--	0.0%	5	1.5%
Unknown	--	0.0%	18	14.8%	18	5.6%

Table 2:

Distribution of patients by periodontal diagnostic classification

Diagnosis	# of Patients
No Periodontitis	70
Periodontitis Stage I - Localized	3
Periodontitis Stage I - Generalized	2
Periodontitis Stage II - Localized	50
Periodontitis Stage II - Generalized	15
Periodontitis Stage III - Localized	91
Periodontitis Stage III - Generalized	59
Periodontitis Stage IV - Localized	12
Periodontitis Stage IV - Generalized	21

Table 3:

Original and refined algorithm performance by disease stage

	De-novo Algorithm				
	Not Perio	**Stage I**	**Stage II**	**Stage III**	**Stage IV**
Sensitivity	72.9%	20.0%	66.2%	74.7%	75.8%
Specificity	81.4%	100.0%	93.4%	87.3%	98.3%
Pos Pred Value	52.0%	100.0%	71.7%	83.6%	83.3%
Neg Pred Value	91.6%	98.8%	91.6%	79.9%	97.3%
	Refined Algorithm				
Sensitivity	90.0%	20.0%	52.3%	87.3%	84.8%
Specificity	85.0%	100.0%	98.5%	90.8%	97.2%
Pos Pred Value	62.4%	100.0%	89.5%	89.1%	77.8%
Neg Pred Value	96.9%	98.8%	89.1%	89.2%	98.3%

De-novo: Overall accuracy = 71.8%; Kappa = 0.59

Refined algorithm: Overall accuracy = 79.6%; Kappa = 0.70

Author Manuscript

Author Manuscript

Author Manuscript

Table 4:

Original and refined algorithm performance by disease stage and grade

	De-novo Algorithm												
	Not Perio	**Stage I: A**	**Stage I: B**	**Stage I: C**	**Stage II: A**	**Stage II: B**	**Stage II: C**	**Stage III: A**	**Stage III: B**	**Stage III: C**	**Stage IV: A**	**Stage IV: B**	**Stage IV: C**
Sensitivity	72.9%	0.0%	50.0%	NA	50.0%	63.0%	0.0%	14.3%	75.9%	26.1%	0.0%	78.9%	30.8%
Specificity	84.4%	100.0%	100.0%	100.0%	97.1%	94.6%	100.0%	99.0%	82.8%	99.3%	100.0%	96.3%	100.0%
PPV	57.3%	NaN	100.0%	NA	25.0%	70.8%	NaN	25.0%	72.1%	75.0%	NaN	57.7%	100.0%
NPV	91.6%	99.4%	99.7%	NA	99.0%	92.5%	99.7%	98.1%	85.4%	94.4%	99.7%	98.6%	97.1%
	Refined Algorithm												
Sensitivity	90.0%	0.0%	50.0%	NA	50.0%	44.4%	0.0%	14.3%	85.3%	26.1%	0.0%	78.9%	30.8%
Specificity	87.3%	100.0%	100.0%	100.0%	98.1%	98.8%	99.7%	99.0%	82.3%	99.3%	100.0%	94.2%	100.0%
PPV	67.0%	NaN	100.0%	NA	33.3%	88.9%	0.0%	25.0%	73.9%	75.0%	NaN	46.9%	100.0%
NPV	96.8%	99.4%	99.7%	NA	99.0%	89.5%	99.7%	98.1%	90.6%	94.4%	99.7%	98.6%	97.1%

De-novo: Overall accuracy = 64.7%; Kappa = 0.54

Refined algorithm: Overall accuracy = 68.8%; Kappa = 0.58

Author Manuscript

Author Manuscript

Author Manuscript

Table 5:

Statistical comparison between original and refined algorithm

	Original	95% CI	Refined	95% CI	test	p-value
STAGE						
Accuracy	0.718	(0.666, 0.767)	0.796	(0.748, 0.838)	−2.290	0.022
Kappa	0.594	(0.524, 0.664)	0.702	(0.640, 0.764)	--	--
STAGE AND GRADE						
Accuracy	0.647	(0.591, 0.699)	0.688	(0.634, 0.739)	−1.101	0.271
Kappa	0.534	(0.467, 0.601)	0.585	(0.521, 0.649)	--	--