

Agreement Among International Periodontal Experts Using the 2017 World Workshop Classification of Periodontitis

Andrea Ravidà¹, Suncica Travan¹, Muhammad H. A. Saleh^{1,2}, Henry Greenwell², Panos N. Papapanou³, Mariano Sanz⁴, Maurizio Tonetti⁵, Hom-Lay Wang¹, Kenneth Kornman¹

¹Department of Periodontics and Oral Medicine, University of Michigan School of Dentistry, Ann Arbor, MI, USA

²Department of Periodontics, University of Louisville School of Dentistry, Louisville, KY, USA.

³Division of Periodontics, Section of Oral, Diagnostic and Rehabilitation Sciences, Columbia University, College of Dental Medicine, New York, NY, USA

⁴Department of Dental Clinical Specialties, ETEP Research Group, Faculty of Odontology, University Complutense of Madrid, Madrid, Spain

⁵Department of Oral and Maxillo-facial Implantology, Shanghai Key Laboratory of Stomatology, National Clinical Research Centre for Stomatology, Shanghai Ninth People Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

Andrea Ravidà and Suncica Travan contributed equally

Corresponding author:

Kenneth Kornman DDS, PhD

University of Michigan School of Dentistry, Department of Periodontics and Oral Medicine

1011 North University Avenue

Ann Arbor, MI 48109

Email: kkornman@umich.edu

Fax: +1(734) 764-2469

Andrea Ravidà DDS, MS

University of Michigan School of Dentistry, Department of Periodontics and Oral Medicine

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/jper.10863](https://doi.org/10.1002/jper.10863).

This article is protected by copyright. All rights reserved.

1011 North University Avenue
Ann Arbor, MI 48109
Email: ravida@umich.edu

Total word count: 3977; Abstract word count: 250
Total number of tables and figures: 2 tables, 2 figures
Number of references: 21; Supplemental Material: 4 Supplementary tables
Key words (MeSH): classification, diagnosis, periodontal diseases, calibration

Running title: Classification of Severe Periodontitis Including Gray Zones

Key Findings: Clinicians trained in 2017 World Workshop periodontitis classification showed moderate concordance in classifying patients with severe periodontitis (stage III and IV).

Abstract:

Background: A new periodontitis classification was recently introduced involving multidimensional staging and grading. The Study's aim was to assess if individuals well-trained in periodontics consistently used the new classification for patients with severe periodontitis. The secondary goal was to identify "gray zones" related to classifications.

Methods: Participants (raters) individually classified 10 pre-selected severe periodontitis cases using the 2017 World Workshop classification. An internet case-based study was conducted after inviting members from American Academy of Periodontology and European Federation of Periodontology. Gold-standard diagnoses were determined by five experts who developed the new periodontitis classification.

This article is protected by copyright. All rights reserved.

Inter-reliability agreement among raters was assessed using Fleiss Kappa index with the jackknife method for linearly-weighted Kappa calculations. McNemar's test was used to determine symmetry between raters and gold-standard panel.

Results: 103 raters participated and classified 9 clinical cases. Fleiss' Kappa values showed moderate inter-examiner consistency among raters for stage (K value: 0.49; 95%CI:0.19-0.79), grade (K value 0.50; 95%CI:0.30-0.70) and extent (K value: 0.51; 95%CI:0.23-0.77). When analyzed as composite (stage, grade, extent) a moderate inter-reliability was present among raters, $k=0.479$ (K value: 0.47; 95%CI: 0.442 – 0.515). Agreement between raters and gold-standard panel was staging 76.6%; grading 82%; and extent 84.8%. In 6 of 9 cases 77-99% of raters consistently agreed with gold-standard panel, and other 3 cases had gray zone factors that reduced rater consistency.

Conclusion: Clinicians trained in the 2017 World Workshop periodontitis classification demonstrated moderate concordance in classifying 9 severe periodontitis cases, and in 6 of 9 cases raters consistently agreed with the gold-standard panel.

Introduction

Periodontitis is a global healthcare problem with increasing costs on individual and societal levels¹, and current evidence indicates that up to 42% of U.S. dentate adults have periodontitis².

The 2017 World Workshop Classification of Periodontal and Peri-Implant Diseases and Conditions agreed on a new periodontal disease classification based on a multifaceted staging and grading system³ that acknowledges a subset of individuals who develop more severe periodontitis than the majority of adults⁴⁻⁶. Furthermore, it has been known for many years that segments of patients with

This article is protected by copyright. All rights reserved.

moderate to severe periodontitis fail to respond predictably to standard periodontitis treatment protocols as administered by well-trained clinicians⁷⁻¹⁰.

The new periodontitis classification system was developed in recognition that prior classifications for periodontal disease did not translate easily into diagnoses and treatment plans for individual patients¹¹. It was acknowledged that there are several periodontitis phenotypes, and the 2017 classification incorporates use of several factors as part of the multifactorial influences on an individual patient's past disease and likelihood of future response to standard principles of periodontal therapy and maintenance^{12, 13}. It remains uncertain whether the structure and integrated information of the staging and grading system allows clinicians to predictably classify individual patients.

The primary goal of this study was to determine the degree of consistency in staging, grading and extent among individuals trained to manage severe periodontitis cases and with prior exposure to the new periodontitis classification^{12, 13}. Cases selected for this study were based on the training and experience of clinicians invited to participate.

Material and Methods:

The study was conducted in compliance with the Helsinki Declaration of 1975, as revised 2013¹⁴. The study was approved by the University of Michigan Medical School Institutional Review Board: Study eResearch ID: HUM00163877, Office for Human Research Protection IRB Registration Number(s): IRB00000246.

This article is protected by copyright. All rights reserved.

Primary and secondary objectives:

The primary objective was to evaluate the inter-examiner agreement among clinicians with experience managing severe periodontitis, i.e. study raters, from North America, United Kingdom, and Europe (Supplementary Table S1); and the agreement among study raters and the gold-standard panel of 5 experts (KK,PNP,MT,MS,HG) who were lead authors of the 2017 World Workshop on periodontitis classification^{12, 13}. The secondary objective was to identify and clarify specific “gray zones” related to subtle aspects of a patient’s classification that may require clarification of specific diagnostic criteria to enhance concordance among raters.

Study Design and Questionnaire:

An internet case-based study was conducted using 10 pre-selected periodontitis cases. Study raters were asked to individually classify each case using stage, grade, and extent of disease based on best interpretation of the 2017 World Workshop Classification criteria¹³. Raters included an international group of clinicians knowledgeable in periodontics including, members of the American Academy of Periodontology (AAP) or the European Federation of Periodontology (EFP). Individuals were invited to participate by email, which contained a link to access the case assessment form.

All raters had prior training on the new classification unrelated to this study, with classification training at various institutions in U.S. and Europe. Additional classification training and calibration were intentionally not provided to raters prior to the study. All raters are listed in Supplementary Table S1.

Participation was voluntary, the responses were anonymous, and study investigators were blinded to raters’ survey responses. Raters were advised of the primary study objectives and were invited to

This article is protected by copyright. All rights reserved.

add comments/questions concerning the selected cases and the classification in general. The timeline for completing the survey was from April 3th to May 8th, 2020. Four reminder follow-up emails (April 8th, 13th, 28th and May 4th) were sent during the survey open period.

Gold-standard diagnosis and qualitative assessment of gray zones

The classification of each case was determined by the 5 gold-standard panel members each of whom independently assessed the 10 cases, submitted their case findings through the online system, and then participated in a series of group discussions during which cases, not diagnosed uniformly by gold-standard panel members, were adjudicated until a unanimous decision was reached by the panel. Raters were invited during the questionnaire to submit comments and questions regarding diagnoses of provided cases, and/or concerning the classification system. Subsequently, critical questions were systematically discussed and answered after closure of the study to generate a consensus statement clarifying each cluster of questions.

Clinical cases:

Ten clinical periodontitis cases were selected by the authors with various disease severities and extent. Clinical and radiographic case information was supplied to raters including: intraoral photos, complete periodontal charts, as well as bitewing and periapical radiographs. Patient age, gender and pertinent medical history were included in the records. Detailed information was provided about medical status and grade modifiers, including smoking and diabetes. All case information provided raters is available (Supplementary Table S2), with removal of identifying information. The published staging and grading tables from the 2017 Workshop were provided as reference for classifying patients¹³. Cases for this study were selected to challenge raters trained to diagnose and manage

This article is protected by copyright. All rights reserved.

severe periodontitis cases. Several cases were intended to reflect staging and grading characteristics that should clearly differentiate Stage II from III, and III from IV, with provided interdental CAL, radiographic bone loss, and tooth loss that should distinguish well-defined stages. Complexities of cases were provided with radiographs and complete periodontal charts. We also sought to include gray zone cases that presented with specific features that posed diagnostic challenges and would likely result in more variance in raters' responses.

Statistical analysis

Power analyses were performed and 100 raters were deemed necessary to estimate the level of variability in the assessment of Stages and Grades among clinicians. It was determined that a minimum of 8 cases had to be classified to ensure that kappa values 0.8 and 0.95 could be detected as significantly different with 80% power and 95% confidence.

The statistical analysis was conducted by an independent biostatistician. Descriptive analyses consisted of absolute and relative frequencies for all categorical parameters. Inter-reliability agreement among raters involved single parameters or matching all three (stage, grade, extent) assessed using Fleiss Kappa index and 95% confidence intervals were calculated using the jackknife

This article is protected by copyright. All rights reserved.

method. Linearly weighted kappa was calculated and McNemar's test was used to evaluate the hypothesis of balanced discordant proportions (symmetry) among raters and the gold-standard panel. Kappa values were interpreted according to Altman-Landis table. The significance level used in the analysis was 5% ($\alpha=0.05$). All analyses were performed using dedicated software^{**‡}.

Footnote:

*R: R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>

† SPSS; Chicago, IL – USA

‡ EPIDAT 4.2 Conselleria de Sanidade, Xunta de Galicia – Spain

Results:

Initially, 164 clinicians well-trained in periodontics were invited to participate as raters. A total of 107 individuals started the survey, but 4 did not complete all case evaluations and were excluded from final analysis. Hence, 103 raters located in USA (68%) or Europe/UK (32%) classified all 10 clinical cases in the study. Of the raters, 35.9% were 30-40 years old, 44.7% between 41-60 and 19.4% over 60. One case (patient #8) was excluded from analysis due to comments from multiple

This article is protected by copyright. All rights reserved.

raters and gold-standard panel members relative to incongruencies between clinical photographs and recorded CAL measurements. Hence, 9 cases were included in the quantitative analysis.

Quantitative Assessment of individual cases and concordance with gold-standard assessment

The relative proportions for each possible answer regarding stage, grade, and extent for every case and the percentage of answers corresponding or diverging from the gold-standard consensus are shown in Figure 1 and Table 1.

The percentage of raters who correctly classified each case according to stage, grade, and extent were calculated. For staging, considering 927 different assessments (103 raters for 9 cases), an inter-examiner overall agreement of 76.6% was found (710 agreements of 927 assessments). Overall, 18.5% of assessments were classified as more severe than the gold-standard ($p < 0.010$; McNemar's test), while 4.8% were classified as less severe. For grading, the overall agreement was 82%, with 760 correct evaluations. Despite substantial concordance, 7.5% of assessments were classified as a significantly lower grade by raters than the gold-standard panel ($p < 0.001$; McNemar's test). Finally, the inter-examiner agreement for extent was 84.8%, with 786 correct evaluations. Raters made 19 extent assessments of "Molar/Incisor patterns" in contrast to the gold-standard panel that diagnosed none. There was a moderate inter-reliability agreement among raters for stage (Fleiss' K value: 0.49; 95%CI: 0.19-0.79), grade (K value 0.50; 95%CI: 0.30-0.70) and extent (K value: 0.51; 95%CI: 0.23-0.77) according to the Altman-Landis scale. When the diagnosis for a single case was analyzed as a composite of three parameters (stage, grade and extent), a moderate inter-reliability was present among raters, $k = 0.479$ (K value: 0.47; 95%CI: 0.442 – 0.515).

This article is protected by copyright. All rights reserved.

The raters' consistency relative to the gold-standard consensus is evident in Figure 2(A-C) and Table 1, where 103 raters consistently scored 6 cases (1,3,4,5,9,10) of 9, with 77-99% agreement with the gold-standard staging (Figure 2A). Staging thresholds were sufficiently clear in 4 of the 6 cases (1,3,4,5) to produce 92-99% agreement among raters and the gold-standard. Stage IV cases, as well as Grade C, had the most consistent diagnosis between raters and the gold-standard.

Raters were independently consistent but less so in some Stage IV cases, such as staging Case #9 (23% rated Stage III) and #10 (16% rated Stage III). Rater agreement with gold-standard grading (Fig 2B) ranged 81-99% for 6 (1,4,5,7,9,10) of 9 cases. Raters consistently scored extent (Fig 2C) for 8 (1,2,3,4,5,7,9,10) of 9 cases ranged from 78-100% that agreed with the gold-standard. See Supplementary Table S4.

For Table 1 Group 2 cases (2,6,7), raters appeared to be less clear in differentiating Stage III/IV because Cases #2 and #7 both have gray zones not clearly defined until recently¹⁵ to guide removal of hopeless teeth to direct Stage IV decisions. Case #6 CAL was defined based on periodontitis but confused some raters based on interproximal CAL influenced by restorative crowns that incorrectly led 71% of raters to assign Stage III rather than the gold-standard consensus Stage II. The interpretation of selected Table 1 cases are discussed in Supplementary Table S3.

The observation that 77-99% of 103 raters independently agreed with the gold-standard panel for 6 of 9 cases. Workshop Classification criteria for staging and grading^{12, 13} translated well for severe

This article is protected by copyright. All rights reserved.

periodontitis cases. However, the challenges that are evident with Cases 2,6,7 indicate that additional clarity of gray zones and boundaries will be necessary in some cases to guide a clinician's focus.

Qualitative assessment of the selected cases: Widespread use of the staging and grading classification system within periodontal training programs has helped identify certain gray zones, that require clinical judgment in the implementation of the new classification¹⁶⁻¹⁸. The present article narratively clarifies certain gray zones of clinical diagnosis to instruct readers on wise use of the new classification. Table 2 shows a step-by-step approach to diagnose each case based on same framework provided by the 2017 Workshop classification^{12, 13}. Case-by-case rationale for diagnoses according to the gold-standard panel is shown in Supplementary Table S4.

DISCUSSION

The new periodontitis classification defines some threshold levels that guide clear decisions for staging and grading patients. As seen in Table 1, the raters' independent staging decisions for four cases strongly agreed (92-99%) with the gold-standard. Since the 2017 classification was released,

This article is protected by copyright. All rights reserved.

multiple reports have provided refinements for implementation of the new classification in daily practice^{15, 16, 19}.

Stages I-IV were designed to classify periodontitis patients based on major distinctions in severity of disease and complexity of treating the patient's periodontal disease and maintaining periodontal health. The stages also generally stratify cases based on need for prior training, experience, and need for multidisciplinary treatment. Grading helps to identify the patient's responsiveness to standard treatment protocols and importance of systemic conditions that may influence or be influenced by periodontitis, which adds further clinician responsibilities. As shown in raters' results above, most cases will be correctly classified if clinicians have basic training in the new classification. This study was designed in part to better understand what gray zones may introduce the need for more specific training in classification.

Gray Zones: Queries from the raters and assessment of the source of variability

The following section describes in detail common gray zones identified via lack of inter-examiner agreement and raters' questions.

1) What are the main determinants used to identify stage? (9 queries)

This article is protected by copyright. All rights reserved.

Stage severity is primarily based on interdental clinical attachment level (CAL), however, only CAL attributable to periodontitis should be used to determine stage. For Case #6, the gold-standard panel determined the most severe interproximal CAL sites were only on teeth with adjacent crowns. The CAL severity that determined Stage II for the case was attributable directly to periodontitis, as opposed to more severe tissue destruction secondary to restorative therapy. However, 71% of raters classified the patient as Stage III based on the most severe interproximal CAL, which was not directly attributable to periodontitis. Although RBL is included as an additional descriptor, both CAL and RBL are assessed at the most affected tooth, with recognition that the CAL resulted from periodontitis.

Suggested reading: (Kornman & Papapanou, 2020; Tonetti, Greenwell, & Kornman, 2018)^{12, 16}

2) *What is considered a periodontally hopeless tooth? Should hopeless teeth be identified and extracted before staging?* (9 queries)

We reaffirm that hopeless teeth can be included when calculating the total number of teeth lost due to periodontitis. As recently published, a hopeless tooth, i.e. one that is considered irrational to treat, is that in which the attachment loss approximates the apex of the root circumferentially in combination with tooth hypermobility (degree 3).

Suggested reading: (Sanz, Papapanou, et al., 2020)¹⁵

3) *Is differentiation between Stage III and IV solely based on complexity factors?* (4 queries)

This article is protected by copyright. All rights reserved.

Stage III and IV require the same essential identifiers, either CAL of ≥ 5 mm or RBL to the middle 1/3 of root or beyond. The gold-standard panel classified Case #7 as Stage III, but 45% of raters classified the case as Stage IV. Stage IV may be differentiated from III either by history of missing teeth attributable to periodontitis plus additional hopeless teeth, as described above¹⁵ to equal ≥ 5 . A Stage IV case should also present with features identified in the classification that define need for complex rehabilitation, including masticatory dysfunction, bite collapse, and potential for losing major dentition components. Case #7 is female age 35 with extensive history of clinical attachment and bone loss. Mobility grade 3 is present in one tooth and grade 2 in two additional teeth. This young patient with generalized severe periodontitis has not lost teeth due to periodontitis and currently has no hopeless teeth that would result in patient's tooth loss ≥ 5 teeth due to periodontitis. In addition, there is minimal masticatory dysfunction or evidence of bite collapse. It is likely that presence of severe bone and connective tissue destruction at early age led 45% of raters to assume patient represents Stage IV disease. As noted by the gold-standard panel, this patient qualifies most clearly as Stage III, indicative of severe periodontitis with potential for additional tooth loss. *Suggested reading: (Tonetti et al., 2018)*¹²

4) Can complexity factors shift a patient whose severity level is compatible with Stage I or II to Stage IV? (2 queries)

Stage IV periodontitis presents either with a level of severity and extent that complex, multidisciplinary rehabilitation, or with periodontitis-associated tooth loss (≥ 5 teeth). Moreover, and in contrast with Stage III disease where severe loss of periodontal tissue support is also prevalent, Stage IV periodontitis jeopardizes a large segment of the dentition or entire dentition,

This article is protected by copyright. All rights reserved.

rather than individual teeth. Consequently, extensive and/or multidisciplinary rehabilitation is required to preserve and restore oral function.

Note that periodontitis cases with severity compatible with Stage I or II cannot be upshifted to Stage IV on the basis of complexity factors alone. Differential diagnosis of Stage IV periodontitis can only be considered in cases where the severity criteria of Stage III periodontitis have been clearly met. It is important to avoid erroneous diagnosis of Stage IV periodontitis in partially edentulous cases with less than 10 occluding pairs, where tooth loss is due to reasons other than periodontitis, such as in cases of primary occlusal trauma, with loss of vertical dimension of occlusion or tooth drifting. Occasionally patients may present with all posterior teeth missing without a known reason for the observed tooth loss, and the clinician may infer the reasons for tooth loss based on patient's oral and general health history and collective assessment of periodontal status of the remaining dentition. The clinician must keep in mind the possibility that prior extraction of all teeth severely affected by periodontitis may result in the survival of only teeth minimally affected by periodontitis, which can distort the accurate staging of the case based on periodontitis-associated tooth loss.

Suggested reading: (Kornman & Papapanou, 2020; Tonetti et al., 2018)^{12, 16}

5) Could more than one extent be assigned for stage? Is there more than one way to calculate the extent of the disease? (2 queries)

A recent publication provides a meaningful method for how the extent of the disease should be calculated¹⁵. The extent of periodontitis is defined by percentage of teeth (not sites) at the stage-defining severity level. The teeth presenting with the specific level of severity (CAL/RBL) used to

This article is protected by copyright. All rights reserved.

assign patient's stage, are the teeth that have to be counted to correctly assess the extent. Depending on whether the number of those teeth is less or more than 30%, the extent is defined as either localized or generalized. *Suggested reading: (Sanz, Papapanou, et al., 2020)*¹⁵

Clinicians who participated as raters were experienced managing severe periodontitis cases and exhibited ability to use the new classification, as described above, consistent with the gold-standard panel. These findings are strongly suggestive that the Table 1 Group 1 cases met Stage III-IV essential CAL and/or radiographic bone loss (RBL) criteria in periodontitis staging table and net tooth loss and/or hopeless teeth, and complexity criteria to qualify clearly for Stage III or IV.

It should be noted that gray zone features may result in non-concordant diagnoses even among knowledgeable clinicians. As seen in Table 1 Group 2, Cases (2,6,7) have same CAL and/or RBL as Group 1, but uncertain numbers of missing/hopeless teeth and lack of clarity for other complexity factors that may guide the assessment of Stage II, III or IV (details in Supplementary Table S3).

Supplementary Staging and Grading Queries

Responses to queries below can be found in the Supplementary Appendix in the online *Journal of Periodontology*. For responses to the queries below, additional information may be found online in the **Supporting Information** section at the end of the article

*When staging a periodontitis patient, which of the missing teeth should be considered periodontal tooth loss? (2 queries). Suggested reading: (Tonetti et al., 2018)*¹²

Should other systemic conditions be considered when grading? (6 queries).

Suggested reading: (Albandar, Susin, & Hughes, 2018; Kornman & Papapanou, 2020)^{16,20}

Is BOP needed for Staging? (5 queries).

*Suggested reading: (Chapple et al., 2018)*²¹

This article is protected by copyright. All rights reserved.

Is there a way to use the new classification for direction to therapy? (2 queries).

Suggested reading: (Sanz et al. 2020)¹⁷

Main findings

This study demonstrated that clinicians with experience in treating periodontitis and prior training in the new periodontitis classification^{12, 13} were capable of independently applying the classification to severe periodontitis cases. Given the multidimensional nature of the classification, some gray zones were expected that provided guidance for certain cases while others are more complex.^{15, 16} Even for cases displaying various complexities, the 103 raters achieved an inter-examiner agreement of 76% for stage, 82% for grade, and 84.8% for extent. As seen in Table 1, certain cases were classified with consistency, ranging from 92-99% of raters and agreement with the gold-standard stage, grade and extent. A second group of patients, Cases #9 and #10, classified stage consistent with the gold-standard Stage IV, yet there was less clarity by raters differentiating Stage III/IV (Case #9: 23% Stage III; Case #10: 16% Stage III). Table 1 Group 2 cases (2,6,7) were not consistently classified by the raters, however Supplementary Table S3 provides interpretation of certain cases that were less clearly defined by currently published classification criteria^{12, 13}.

This article is protected by copyright. All rights reserved.

This study has limitations including intentional absence of incipient periodontitis cases and only one Stage II case of moderate periodontitis, designed to target this study's primary clinician audience focused on managing complex periodontal disease cases. Future studies evaluating inter-examiner agreement using the new classification among raters with a broader patient range of periodontal disease and experience would likely benefit from inclusion of additional Stage I and II cases that may clarify boundaries across the full range of staging and grading to be experienced within the dental community.

In addition, understanding the educational training required for a broader range of clinical exposure necessary to integrate the new classification into periodontal disease management by a wide range of general dentists and specialists. As noted above we selected some cases that introduced gray zones expected to complicate the consistency of the classification among well-trained but diverse range of clinicians.

In terms of grading, the present study demonstrates that agreement among raters and consistency with the gold-standard, is derived from substantial variability in radiographic bone levels, as was evident from Figure 2. Among selected cases, #2 and #6 displayed greater variability. The key to a correct grading of most cases in the absence of direct evidence of progression is the correct measurement of percentage bone loss/age. For Case #2 (Grade C), at least ≥ 1 tooth (lower lateral incisors) has RBL of $>68\%$ (1.0 ratio to patient's age) consistent with Grade C, but 41% of raters

This article is protected by copyright. All rights reserved.

underestimated and selected Grade B. For Case #6 (Grade B), none of the teeth have RBL of $\geq 52\%$ (1.0 ratio to patient's age).

In the present study, a broad group of clinicians with expertise in periodontics have been asked to apply the new classification to cases of periodontitis to practically investigate the way in which the new classification is interpreted and utilized broadly. Selecting this group of clinicians was made with awareness of their key role as educators, and that their calibration is key to correctly and reliably training dental providers. In addition, one of the objectives of the study was to identify gray zones that were not specifically defined in the Workshop papers on classification of periodontitis^{12, 13}.

Cases for the study were not intended to be representative of the general population but were selected to represent severe periodontitis cases as defined in the 2017 Workshop papers^{12, 13}. The provided percentages and coefficients of agreement are based on a small sample of cases selected because of diagnostically challenging presentations, which likely played a crucial role for the relatively moderate kappa of 0.5 inter-reliability. The agreement rate on diagnosis of periodontitis might have been different if raters and experts had been asked to classify a larger random sample of patients; however, as shown in Figure 1 and Table 1, this study intentionally enrolled severe cases to determine consistency of the 103 raters who were 77-99% in agreement for 6 of 9 cases.

This study and others are designed to educate current and future dental providers in comprehension of the new classification through continuous training and assessment of gray zones. For this reason,

This article is protected by copyright. All rights reserved.

we will perform similar studies with the aim to further increase the consistency observed for the diagnosis of periodontal patients that must consider gray zones to accurately classify the patients. Training a broad range of clinicians in the interpretation of specific gray zones required reliably differentiating Stages II/III and III/IV.

CONCLUSIONS

Clinicians appropriately trained in the management of periodontal disease patients demonstrated an overall moderate concordance in their diagnostic classification of periodontitis cases utilizing the 2017 World Workshop Classification system. Although severe periodontitis cases displaying well-defined features were consistently rated, gray zones that introduced interpretational complexities decreased the inter-rater agreement. Continuous identification, refinement, and incorporation of such diagnostic challenges are needed to guide the training of a broad range of clinicians and predictably enhance their diagnostic agreement in the classification of periodontitis.

ACKNOWLEDGMENTS

The authors thank all the raters (Supplementary Table S1) for their commitment and essential role in completing the survey, and Adam Barragato (University of Michigan) for helping to construct the survey. Furthermore, the authors thank Drs. Giorgio Pagni (University of Milan), Erin Shan-Huey Yu (ESY, University of Michigan), Giacomo Gualini (University of Turin) and Marco Clementini (San Raffaele University) for providing some challenging clinical cases. Finally, the authors want to thank Juan Luis Gomez (stHalley statistics Barcelona) for analysis of the data. This manuscript was supported in part by the University of Michigan Periodontal Graduate Student Research Fund. Competing Interests: The authors report no conflicts of interest related to this study.

This article is protected by copyright. All rights reserved.

AUTHOR CONTRIBUTIONS

Conception and design: AR,ST,MHS,KK; Provided clinical cases: AR,MHS, and others listed in Acknowledgements; Selection of cases: AR,MHS,ST,KK; Invitation of clinicians to serve as raters: AR,ST,KK; Infrastructure to allow anonymous independent assessment of cases: ST,AR;

Statistical plan, analysis, and interpretation of the data: AR, MHS, ST, KK; Consensus development

Among authors: MS, KK,HG,MT,PNP; Drafting article: AR,MHS,ST,MS,KK,HG,MT,PNP;

Collection/assembly of data: AR,MHS,ST; Critical revision for important intellectual content: KK,HG,MT,PNP,MS,HLW; Approval of the revised article for publication: AR,ST,MS,MHS, PNP,HG,MT,HLW,KK

References:

1. Tonetti MS, Jepsen S, Jin L, Otomo-Corgel J. Impact of the global burden of periodontal diseases on health, nutrition and wellbeing of mankind: A call for global action. *J Clin Periodontol* 2017;44:456-462.
2. Eke PI, Thornton-Evans GO, Wei L, Borgnakke WS, Dye BA, Genco RJ. Periodontitis in US Adults: National Health and Nutrition Examination Survey 2009-2014. *J Am Dent Assoc* 2018;149:576-588 e576.
3. 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.
4. Billings M, Holtfreter B, Papapanou PN, Mitnik GL, Kocher T, Dye BA. Age-dependent distribution of periodontitis in two countries: Findings from NHANES 2009 to 2014 and SHIP-TREND 2008 to 2012. *J Periodontol* 2018;89 Suppl 1:S140-S158.
5. Loe H, Anerud A, Boysen H, Morrison E. Natural history of periodontal disease in man. Rapid, moderate and no loss of attachment in Sri Lankan laborers 14 to 46 years of age. *J Clin Periodontol* 1986;13:431-445.

This article is protected by copyright. All rights reserved.

6. Baelum V, Fejerskov O, Karring T. Oral hygiene, gingivitis and periodontal breakdown in adult Tanzanians. *J Perio Res* 1986;21:221-232.
7. Hirschfeld L, Wasserman B. A long-term survey of tooth loss in 600 treated periodontal patients. *J Periodontol* 1978;49:225-237.
8. McFall WT, Jr. Tooth loss in 100 treated patients with periodontal disease. A long-term study. *J Periodontol* 1982;53:539-549.
9. Linde J, Nyman S. Long-term maintenance of patients treated for advanced periodontal disease. *J Clin Periodontol* 1984;11:504-514.
10. McGuire MK, Nunn ME. Prognosis versus actual outcome. IV. The effectiveness of clinical parameters and IL-1 genotype in accurately predicting prognoses and tooth survival. *J Periodontol* 1999;70:49-56.
11. Armitage GC, Cullinan MP. Comparison of the clinical features of chronic and aggressive periodontitis. *Periodontol 2000* 2010;53:12-27.
12. 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-S172.
13. 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-S182.
14. World Medical Association. Declaration of Helsinki: Ethical principles for medical research involving human subjects. *Journal American Med Assoc* 2013;310(20): 2191-2194.
15. Sanz M, Papapanou PN, Tonetti MS, Greenwell H, Kornman K. Guest editorial: Clarifications on the use of the new classification of periodontitis. *J Periodontol* 2020;91:1385.
16. Kornman KS, Papapanou PN. Clinical application of the new classification of periodontal diseases: Ground rules, clarifications and "gray zones". *J Periodontol* 2020;91:352-360.
17. Sanz M, Herrera D, Kebschull M, et al. Treatment of stage I-III periodontitis-The EFP S3 level clinical practice guideline. *J Clin Periodontol* 2020;47 Suppl 22:4-60.
18. Ravida A, Qazi M, Troiano G, et al. Using periodontal staging and grading system as a prognostic factor for future tooth loss: A long-term retrospective study. *J Periodontol* 2020;91:454-461.

This article is protected by copyright. All rights reserved.

19. Tonetti MS, Sanz M. Implementation of the new classification of periodontal diseases: Decision-making algorithms for clinical practice and education. *J Clin Periodontol* 2019;46:398-405.
20. Albandar JM, Susin C, Hughes FJ. Manifestations of systemic diseases and conditions that affect the periodontal attachment apparatus: Case definitions and diagnostic considerations. *J Periodontol* 2018;89 Suppl 1:S183-S203.
21. Chapple ILC, 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. *J Periodontol* 2018;89 Suppl 1:S74-S84.

SUPPORTING INFORMATION

Additional supporting information: ***Supplementary Staging and Grading Queries***

may be found online in the Supporting Information section at the end of the article.

Figure Legends

Figure 1: Relative proportions for each possible answer regarding stage, grade, and extent for every case and the percentage of answers corresponding to or diverging from the gold-standard.

Figure 2: The height of each bar indicates the percentage of 103 raters who correctly classified each of 9 cases according to gold-standard consensus for stage 2(A), grade 2(B), and extent 2(C). Cases were identified on horizontal axis by a number e.g. (1) and gold-standard consensus for Stage III (1), Grade B (1), and Extent Gen (1).

This article is protected by copyright. All rights reserved.

Table 1

| | Stage Case Group 1 | | | Grade Case Group 1 | | | Extent Case Group 1 | | |
|----------------------------|--------------------|--------------|-------------|--------------------|--------------|-------------|---------------------|--------------|--------------------------|
| | Gold-std Stage | Rater % | Rater Stage | Gold-std Grade | Rater % | Rater Grade | Gold-std Extent | Rater % | Rater Extent |
| Case #1 | III | 98% | III | B | 81% | B | Generalized | 88% | Generalized |
| Case #3 | III | 92% | III | B | 67% 25% | B C | Generalized | 80% | Generalized |
| Case #4 | IV | 99% | IV | C | 93% | C | Generalized | 99% | Generalized |
| Case #5 | III | 95% | III | C | 91% | C | Localized | 78% | Localized |
| Case #9 | IV | 77% | IV | C | 99% | C | Generalized | 97% | Generalized |
| Case #10 | IV | 84% | IV | C | 98% | C | Generalized | 100% | Generalized |
| % Average Agreement | | 90.8% | | | 88.2% | | | 90.2% | |
| | | | | | | | | | |
| | Stage Case Group 2 | | | Grade Case Group 2 | | | Extent Case Group 2 | | |
| Case #2 | III | 63% 36% | III IV | C | 54% 41% | C B | Generalized | 100% | Generalized |
| Case #6 | II | 27% 71% | II III | B | 56% 40% | B A | Generalized | 28% 72% | Generalized Localized |
| Case #7 | III | 54% 45% | III IV | C | 98% | C | Generalized | 93% | Generalized |
| % Average Agreement | | 48.0% | | | 69.3% | | | 73.7% | |
| | | | | | | | | | |

Table 1. The nine clinical cases included in the quantitative analysis are stratified by Group 1 and Group 2 based on percentage of the 103 raters that independently agreed with the stage determined by the five person gold-standard panel as shown on the left. Group 1 includes six cases for which $\geq 75\%$ of raters agreed with the gold-standard Stage for each Group 1 case. Four of the six Group 1 cases had rater agreements that ranged between 92-99%. Group 2 included three cases for which $< 75\%$ of raters agreed with the gold-standard panel, and the remainder of raters' decisions

This article is protected by copyright. All rights reserved.

that differed from the gold-standard Stage are also shown. For each case the gold-standard Grade and Extent are shown with the raters' Grade and Extent.

Table 2

Table 2

| | | Case 1 | Case 2 | Case 3 | Case 4 | Case 5 | Case 6* | Case 7 | Case 9 | Case 10 |
|------------------------------|---|--|--|--|--|---|--|---|---|---|
| Stage | Interdental CAL at the site of greatest loss | ≥1 site has CAL of ≥ 5 mm | ≥1 site has CAL of ≥ 5 mm | ≥1 site has CAL of ≥ 5 mm | ≥1 site has CAL of ≥ 5 mm | ≥1 site has CAL of ≥ 5 mm | ≥1 site has CAL of ≥ 5 mm | ≥1 site has CAL of ≥ 5 mm | ≥1 site has CAL of ≥ 5 mm | ≥1 site has CAL of ≥ 5 mm |
| | RBL | (UR 1st premolar) has RBL of ≥ 33% | ≥ 1 tooth has RBL of ≥ 33% | ≥ 1 tooth has RBL of ≥ 33% | ≥ 1 tooth has RBL of ≥ 33% | ≥ 1 tooth has RBL of ≥ 33% | ≥ 1 tooth has RBL of ≥ 15%. None of the teeth has RBL ≥ 33%. | ≥ 1 tooth has RBL of ≥ 33% | ≥ 1 tooth has RBL of ≥ 33% | ≥ 1 tooth has RBL of ≥ 33% |
| | Tooth loss due to periodontitis | N/A ^a | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| | Complexity factors | PD ≥ 6mm Furcation CL II Vertical ≥ 3 mm | PD ≥ 6mm Furcation CL II, III | PD ≥ 6mm Furcation CL II | Need for complex rehabilitation. Less than 20 teeth present. | PD ≥ 6mm Vertical ≥ 3 mm | — | PD ≥ 6mm Furcation CL II Vertical ≥ 3 mm | Generalized mobility ≥ grade 2 | Generalized mobility ≥ grade 2 Need for complex rehabilitation. Severe ridge defects. |
| Gold standard Opinion | | III | III | III | IV | III | II | III | IV | IV |
| Extent^o | | Percentage of teeth with CAL ≥ 5mm is more than 30%. | Percentage of teeth with CAL ≥ 5mm is more than 30%. | Percentage of teeth with CAL ≥ 5mm is more than 30%. | Percentage of teeth with CAL ≥ 5mm is more than 30%. | Percentage of teeth with CAL ≥ 5mm is less than 30%. | Percentage of teeth with CAL ≥ 5mm is more than 30%. | Percentage of teeth with CAL ≥ 5mm is more than 30%. | Percentage of teeth with CAL ≥ 5mm is more than 30%. | Percentage of teeth with CAL ≥ 5mm is more than 30%. |
| Gold standard Opinion | | Generalized | Generalized | Generalized | Generalized | Localized | Generalized | Generalized | Generalized | Generalized |
| Grade | Longitudinal Data | — | — | — | — | — | — | — | — | — |
| | Rate of RBL/age | ≥ 1 tooth has RBL of > 14% (0.25 ratio to patient's age). None of the teeth has RBL of > 56% (1.0 ratio to patient's age) | ≥ 1 tooth (lower lateral incisors) has RBL of > 68% (1.0 ratio to patient's age) | ≥ 1 tooth has RBL of > 15% (0.25 ratio to patient's age). None of the teeth has RBL of ≥ 59% (1.0 ratio to patient's age) | ≥ 1 tooth has RBL of > 15% (0.25 ratio to patient's age). None of the teeth has RBL of ≥ 59% (1.0 ratio to patient's age) | ≥ 1 tooth (UR second molar) has RBL ≥ 45% (1.0 ratio to patient's age). | ≥ 1 tooth has RBL of > 13% (0.25 ratio to patient's age). None of the teeth has RBL of ≥ 52% (1.0 ratio to patient's age) | ≥ 1 tooth has RBL of > 35% (1.0 ratio to patient's age) | ≥ 1 tooth has RBL of ≥ 27% (1.0 ratio to patient's age) | ≥ 1 tooth has RBL of ≥ 47% (1.0 ratio to patient's age) |
| | Case Phenotype | — | — | — | — | — | — | — | — | — |
| | Smoking | — | — | — | Smoking ≥ 10 cigarettes/day | — | — | — | — | Smoking ≥ 10 cigarettes/day |
| | Diabetes | — | — | — | A1C= 5.6%. | A1C= 10.4%. | — | — | — | — |
| | High sensitivity CRP | — | — | — | — | — | — | — | — | — |
| Gold standard Opinion | | B | C | B | C | C | B | C | C | C |

* Identifying Stage involves assessing clinical attachment loss, but also involves clinical judgement. Clinicians should not ignore the clinical parameters per given case yet evaluate the case in a holistic approach.

This article is protected by copyright. All rights reserved.

Good clinical judgment is paramount to arrive at a diagnosis that best represents the totality of the clinical situation in a holistic approach. When the severity of the clinical parameters is borderline, the degree of case complexity should be re-assessed.

Case 6 exemplifies a border-line situation. While the pattern of bone loss radiographically is horizontal and the severity of bone loss does not extend beyond the coronal third of the root length, there are a couple of sites with CAL = 5/6 mm. Assuming CALs were accurate in this case, the loss of attachment is possibly contributed by local factors (crowns), not due to advanced disease. The new classification defined Stage III as “severe periodontitis with potential for additional tooth loss”. We do not see this case to be presented with such disease severity. The fact that staging and grading are patient-based, not tooth-based concepts cannot be overstated. While staging this case as III will not be necessarily “wrong”, a Stage II for this case is more in line with the “spirit” of the new classification. At least from a management standpoint, we do not see this case as complex per se. Although we stand by CAL as the primary stage determiner, this case exhibits an example where RBL may give a more well-rounded diagnosis.

°As defined by the percentage of teeth at the stage-defining severity level (Sanz et al., 2020)¹⁵

¶ The history and number of teeth lost due to periodontitis were not provided.

CAL: Clinical attachment level; RBL: Radiographic bone loss; UR: Upper right.

| | Stage Case Group 1 | | | Grade Case Group 1 | |
|----------------------------|--------------------|---------|-------------|--------------------|---------|
| | Gold-std Stage | Rater % | Rater Stage | Gold-std Grade | Rater % |
| Case #1 | III | 98% | III | B | 81% |
| Case #3 | III | 92% | III | B | 67% |
| | | | | | 25% |
| Case #4 | IV | 99% | IV | c | 93% |
| Case #5 | III | 95% | III | c | 91% |
| Case #9 | IV | 77% | IV | c | 99% |
| Case #10 | IV | 84% | IV | c | 98% |
| % Average Agreement | 90.8% | | | 88.2% | |

This article is protected by copyright. All rights reserved.

Case 2

Case 3

Case 4

Case 5

Case 6*

Manuscript

| | Stage Case Group 2 | | | Grade Case Group | |
|----------------------------|--------------------|------|-----|------------------|-------|
| Case #2 | III | 63% | III | c | 54% |
| | | 36% | IV | | 41% |
| Case #6 | III | 27% | II | B | 56% |
| | | 71% | III | | 40% |
| Case #7 | III | 54% | III | c | 98% |
| | | 45% | IV | | |
| % Average Agreement | | 48.% | | | 69.3% |

Aut

This article is protected by copyright. All rights reserved.

| | | | | | |
|---|--|--|--|--|--|
| ≥1 site has CAL of ≥ 5 mm | ≥1 site has CAL of ≥ 5 mm | ≥1 site has CAL of ≥ 5 mm | ≥1 site has CAL of ≥ 5 mm | ≥1 site has CAL of ≥ 5 mm | ≥1 site has CAL of ≥ 5 mm |
| ≥ 1 tooth has RBL of ≥ 33% | ≥ 1 tooth has RBL of ≥ 33% | ≥ 1 tooth has RBL of ≥ 33% | ≥ 1 tooth has RBL of ≥ 33% | ≥ 1 tooth has RBL of ≥ 15%. <u>None of the teeth has RBL ≥ 33%.</u> | ≥ 1 tooth has RBL of ≥ 15%. |
| N/A | N/A | N/A | N/A | N/A | N/A |
| PD ≥ 6mm Furcation CL II. III | PD ≥ 6mm Furcation CL II | Need for complex rehabilitation. Less than 20 teeth present. | PD ≥ 6mm Vertical ≥ 3 mm | — | — |
| III | III | IV | III | II | — |
| Percentage of teeth with CAL ≥ 5mm is more than 30%. | Percentage of teeth with CAL ≥ 5mm is more than 30%. | Percentage of teeth with CAL ≥ 5mm is more than 30%. | Percentage of teeth with CAL ≥ 5mm is less than 30%. | Percentage of teeth with CAL ≥ 5mm is more than 30%. | Percentage of teeth with CAL ≥ 5mm is more than 30%. |
| Generalized | Generalized | Generalized | Localized | Generalized | Generalized |
| — | — | — | — | — | — |
| ≥1 tooth (lower lateral incisors) has RBL of > 68% (1.0 ratio to patient's age) | ≥ 1 tooth has RBL of > 15% (0.25 ratio to patient's age). None of the teeth has RBL of ≥ 59% (1.0 ratio to patient's age) | ≥ 1 tooth has RBL of > 15% (0.25 ratio to patient's age). None of the teeth has RBL of ≥ 59% (1.0 ratio to patient's age) | ≥1 tooth (UR second molar) has RBL ≥ 45% (1.0 ratio to patient's age). | ≥ 1 tooth has RBL of > 13% (0.25 ratio to patient's age). None of the teeth has RBL of ≥ 52% (1.0 ratio to patient's age) | ≥ 1 tooth has RBL of > 13% (0.25 ratio to patient's age). None of the teeth has RBL of ≥ 52% (1.0 ratio to patient's age) |

This article is protected by copyright. All rights reserved.

Table 2

| | | | | | |
|---|---|-------------------------------------|-------------|---|--|
| | | | | | |
| — | — | — | — | — | |
| — | — | Smoking ≥ 10 cigarettes/day | — | — | |
| — | — | A1C= 5.6%. | A1C= 10.4%. | — | |
| — | — | — | — | — | |
| C | B | C | C | B | |