ORIGINAL ARTICLE

Intra- and inter-examiner reliability in classifying periodontitis according to the 2018 classification of periodontal diseases

Lory Abrahamian[1] | Andrés Pascual-LaRocca[1] | Lucía Barallat[1] | Cristina Valles[1] | David Herrera[2] | Mariano Sanz[2] | José Nart[1] | Elena Figuero[2]

[1]Department of Periodontology, Universitat Internacional de Cataluña, Barcelona

[2]ETEP (Etiology and Therapy of Periodontal and Peri-Implant Diseases) Research Group, Complutense University of Madrid, Madrid, Spain

Correspondence
Lucía Barallat, Department of Periodontology, Universidad Internacional de Cataluña, C/Josep Trueta s/n, Sant Cugat del Vallès, Barcelona 08195, Spain.
Email: lbarallat@uic.es

Abstract

Aim: To evaluate the inter-examiner reliability in classifying periodontitis using the 2018 classification of periodontal diseases, when used by postgraduate students, academics, and specialist clinicians trained in European Federation of Periodontology (EFP) and American Academy of Periodontology (AAP) postgraduate-accredited programmes.

Materials and Methods: An online survey including five patients with periodontitis was sent twice to seven specialists in periodontology to provide the staging and grading characteristics. After agreeing on a "gold-standard" classification, the same questionnaire was sent to 16 EFP and 73 AAP postgraduate programmes, to be answered by their faculty, graduates, and students. The responses were compared with the gold-standard classification, and the inter-examiner agreement was calculated.

Results: One-hundred and seventy-four participants completed the survey. The inter-examiner agreement resulted in 68.7% in assigning the stage, 82.4% in assigning the grade, and 75.5% in assigning the extent. The academic position and the experience of the participants did not have any significant influence on classifying periodontitis as the gold standard.

Conclusions: The use of the 2018 periodontitis classification resulted in high inter-examiner reliability when used by a specialist group of clinicians, postgraduate students, and academicians, irrespective of their current position and experience. Given the low response rate and potential selection bias, results pertaining to the use of this system in classifying periodontitis should be interpreted with caution.

KEYWORDS
classification, diagnosis, periodontal diseases, periodontitis, survey

Clinical Relevance

Scientific rationale for study: A new periodontitis classification scheme was adopted during the 2018 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions, in which "periodontitis" is further characterized based on a multi-dimensional staging and grading system. There is a need to assess the reliability of this classification.

Principal findings: The 2018 classification of periodontitis has a very high inter-examiner reliability in a specialist group of postgraduate students and periodontists.

Practical implications: These results suggest that this new classification can be used accurately to classify periodontitis. However, given the potential low response rate and selection bias, caution is needed in interpreting results.

1 | INTRODUCTION

Classification systems not only help to provide frameworks that permit studying the aetiology and pathogenesis of diseases but also support the healthcare community by communicating in a common language and serve as a starting point to arrive at a patient-centred diagnosis (Armitage, 2014).

Classifications of periodontal diseases have been repeatedly modified from its first international recognition in 1942 (Orban, 1942) until 2018 (Caton et al., 2018) in an attempt to align it with emerging scientific evidence. Researchers have introduced various case definitions for periodontal diseases based on etiologic factors, pathologic changes, or clinical manifestations. Since the previous internationally accepted classification system published in 1999 (Armitage, 1999), substantial new information has emerged from population studies, basic science investigations, and the evidence from prospective studies evaluating environmental and systemic risk factors. The analysis of this evidence prompted the 2018 World Workshop organized by the European Federation of Periodontology (EFP) and the American Academy of Periodontology (AAP) to develop a new classification framework for periodontal and peri-implant diseases and conditions, including periodontitis.

According to this new periodontal disease classification scheme, forms of periodontitis previously recognized as "chronic" or "aggressive" were now grouped under a single category "periodontitis" and were further characterized using a multi-dimensional staging and grading system (Papapanou et al., 2018). Staging mainly depends 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 periodontitis progression, assessment of the risk for further progression, analysis of possible poor outcomes of treatment, and assessment of the risk factors that may influence the disease or its treatment (Papapanou et al., 2018; Tonetti et al., 2018). The aim of this staging and grading system was to guide clinicians in the treatment planning of patients with periodontitis (Sanz, Herrera, et al., 2020; Sanz, Papapanou, et al., 2020) and to support them in detecting patients with a high risk of disease progression and/or who are less likely to respond predictably to standard periodontal treatment (Kornman & Papapanou, 2020).

This new classification system, which differs considerably from the previous one, may constitute a challenge in the process by which periodontists/dentists usually formulate their diagnoses (Graetz et al., 2019) and may confuse practitioners when relating the new nomenclature to the clinical diagnosis of their patients (Milward & Chapple, 2003). Therefore, the primary objective of this observational study was to assess the inter-examiner reliability of the new classification of periodontitis among a specialist group comprised of university faculty, specialist clinicians, and postgraduate students.

2 | MATERIALS AND METHODS

2.1 | Study design

This observational cross-sectional study was designed following the STARD guidelines (Standards for Reporting of Diagnostic Accuracy, Cohen et al., 2016) since it evaluates the use of a new classification system (Caton et al., 2018) as a diagnostic tool. Ethical approval for the study was obtained from the Scientific Committee of the Universitat Internacional de Catalunya (UIC) (Barcelona, Spain) (PER-ENC-2018-02).

This study was based on the examination of the baseline digital documentation and subsequent stage, extent, and grade definition of five untreated periodontitis cases, presented in the form of an online survey.

2.2 | Survey

Five periodontitis cases from the archive of patients of the Periodontology Department at the UIC (Barcelona, Spain) were randomly selected using a randomization software from a database of 30 patients undergoing periodontal treatment. These patients had provided informed consent to the use of their data, which were anonymized, in the context of training and research. Gingival diseases, periodontitis as manifestation of systemic diseases, acute periodontal lesions, and presence of dental implants were considered as exclusion criteria. The case description included a general outline of the patient's medical and dental history, intra-oral photographs, a panoramic radiograph, a full set of periapical radiographs, and periodontal charting with the following clinical periodontal measures: probing depth, plaque scores (visually evaluated after the use of a disclosing solution, as present or absent), bleeding on probing, clinical attachment loss (CAL), tooth mobility (Miller, 1985), and furcation involvement (Hamp, 1975). The medical history, specifying information about relevant medical aspects, such as glycaemic control and tobacco use, was also provided. Figure 1 shows a representative example of one of the cases. For specific details regarding the five cases, please see the case presentations in Supporting Information. Prior to starting the study, all the probands were informed on the details of the study and agreed to participate by signing an online informed consent. The participants were asked to evaluate each case independently and to provide a classification (stage, grade, and extent of periodontitis) using the "2018 Periodontitis Classification of
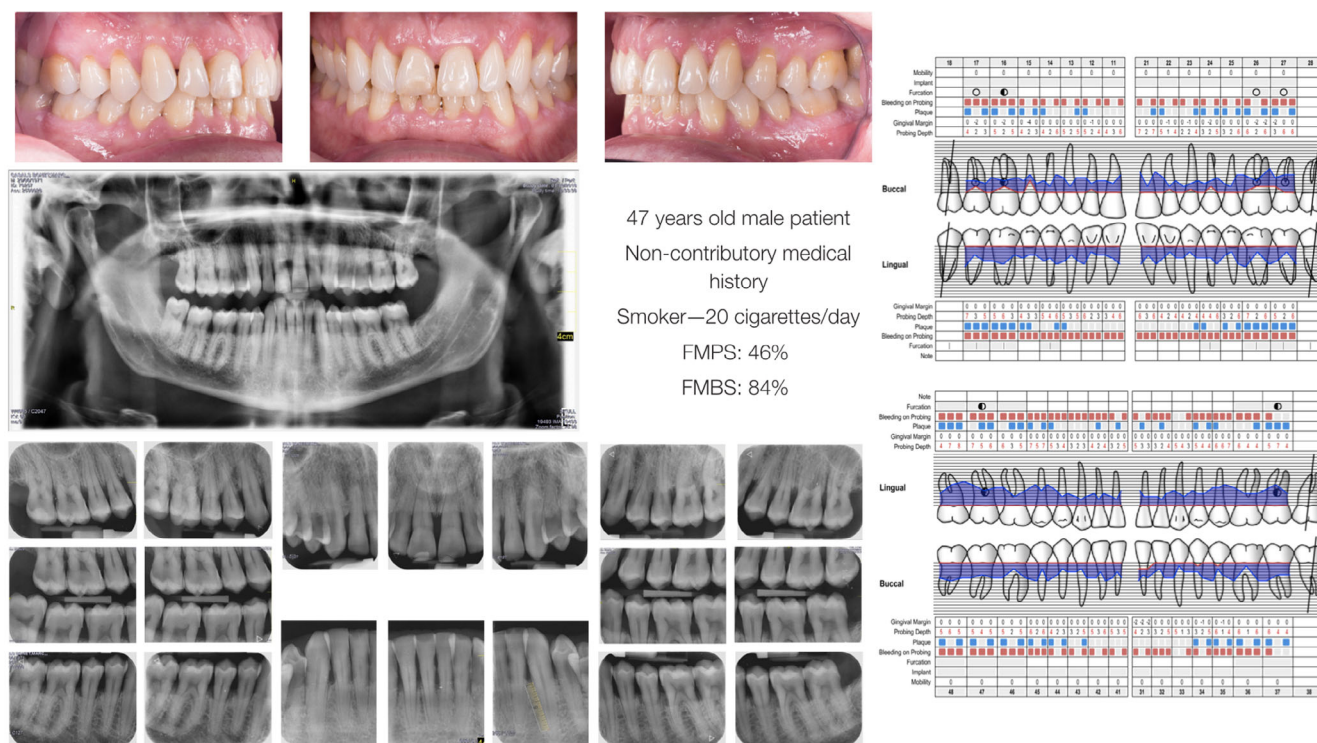
FIGURE 1 Representative example of a case in the survey. FMBS, full-mouth bleeding score; FMPS, full-mouth plaque score

the World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions" (Caton et al., 2018; Papapanou et al., 2018; Tonetti et al., 2018), following the associated algorithm developed by Tonetti and Sanz (2019) and by responding to close-ended questions. The online survey documents including the five cases were created in English using the Google Forms platform from April to May 2020. Once the survey was completed, the answers were saved and visible to a single examiner (Lory Abrahamian).

2.2.1 | Experts' evaluation

The first step included the evaluation of the survey by seven internationally recognized experts in the field from the UIC and from the University Complutense of Madrid (José Nart, Cristina Valles, Andrés Pascual, Lucía Barallat, Mariano Sanz, David Herrera, and Elena Figuero). The survey including the five cases was sent twice, with a minimum timespan of 7 days from the first classification. Then, in June 2020, an agreement on the periodontal classification on each case was set up among these experts by open discussion in videoconference where the final reference classification was established and considered as the gold-standard classification for the second part of the study.

2.2.2 | General survey

In the second phase, the same survey was sent to 16 EFP- and 73 AAP-accredited postgraduate programmes. The link to access electronically the survey was sent to the programme directors, who were asked to forward it to their faculty, graduates, and postgraduate students. Respondents were then categorized according to their academic position and experience into postgraduate students, specialist clinicians, and university faculty. Postgraduate students were considered dentists currently enrolled in the periodontology masters of the accredited programmes, while specialist clinicians were considered board-certified periodontists who are dedicated to private practice and are not involved in academics. University faculty were considered as periodontists currently teaching in the accredited programmes; they could also sometimes be alumni of the respected specialist programme but not necessarily.

2.3 | Statistical analysis

The primary outcome variable was the agreement of the staging, grading, and extent with the established gold-standard classification. Secondary outcomes, considered as potential explanatory outcomes, were the years of experience and the academic position.

In the first part of the study, the intra-examiner reliability of the seven experts developing the gold standard was evaluated by calculating the percentage of concordance and the kappa score. This was calculated at the first and second classifications of any expert, without any distinction between experts. Subsequently, the intra-examiner agreement was attained by comparing any expert's response between the first and second "diagnoses." For the extent, unweighted kappa scores were calculated, while for the stage and

TABLE 1 Intra-examiner agreement of the seven experts

	Concordance (%)	Kappa	95% CI	p-Value
Stage	82.30	0.71	(0.48–0.93)	<.01
Grade	91.40	0.85	(0.71–0.99)	<.01
Extent	83.00	0.52	(0.17–0.87)	.01

Note: Unweighted kappa for extent and weighted kappa for stage and grade.
Abbreviation: CI, coefficient interval.

TABLE 2 Gold-standard classification

	Stage	Grade	Extent
Case 1	III	C	Generalized
Case 2	IV	C	Generalized
Case 3	III	B	Generalized
Case 4	III	C	Generalized
Case 5	III	B	Localized

TABLE 3 Sociodemographic data of the sample

	n (%)
Gender	
Male	101 (58.7)
Female	71 (41.3)
Current position	
University faculty	77 (44.2)
Specialist clinician	36 (20.7)
Postgraduate student	61 (35.1)
Experience (years)	
0	37 (21.6)
<5	43 (25.1)
5–10	30 (17.5)
>10	61 (35.7)

Note: Absolute frequency and percentage: *n* (%).

grade, weighted kappa scores were evaluated (Fleiss, 1981). A six-level nomenclature was used to interpret the kappa values: poor agreement = <0.00; slight agreement = 0.00–0.20; fair agreement = 0.21–0.40; moderate agreement = 0.41–0.60; substantial agreement = 0.61–0.80, and almost perfect agreement = 0.81–1.00 (Landis & Koch, 1977). These results were used to agree on a "gold-standard" classification.

In the second part of the study, descriptive data were presented as absolute frequencies and percentages (%). The association between agreement and potential explanatory outcomes was analysed using chi-square tests and a logistic regression model. Two multivariate logistic regression models were constructed for agreement of stage, grade, and extent as dependent variables, and current position and years of experience as independent variables. The results of the models were reported as adjusted odds ratios (ORs) and 95% confidence interval (CI). Reference categories were determined as "specialist clinician" and "<5 years of experience."

Sample size calculation resulted in the estimation of 90 participants, assuming an expected 70% agreement, which was considered as substantial (α risk = 5%, β risk = 10%) in a bilateral contrast, and a response rate of 30%. The level of significance was set at .05. The version 3.5.2 of the software R was used for all analyses.

3 | RESULTS

3.1 | Experts

3.1.1 | Intra-examiner agreement

The intra-examiner reliability by the experts' evaluation resulted in 82.30% concordance in the stage (kappa score = 0.71, 95% CI [0.48–0.93]; p < .001), 91.40% concordance in the grade (kappa score = 0.85, 95% CI [0.71–0.99]; p < .001) and 83% concordance in the extent (kappa score = 0.52, 95% CI [0.17–0.87]; p = .001) (Table 1).

3.1.2 | Assessment of gold-standard classification

The resulting "gold-standard" classification corresponded to four cases defined as stage III and one case as stage IV; four cases were generalized while one was localized (Table 2).

3.2 | General survey

The "Periodontitis Cases Online Survey" was completed by 174 participants, 58.7% being male and 39.4%, between 30 and 39 years of age. The sociodemographic characteristics of the sample are presented in Table 3.

3.2.1 | Inter-examiner agreement

The comparison of the participants' responses to the gold standard resulted in an overall percentage of agreement of the stage of 68.7%, the grade of 82.4%, and the extent of 75.5%.

Neither the current academic position nor the experience of the participants had a statistically significant influence on the level of agreement (p > .05). Table 4 shows the absolute frequencies and percentages of agreement of the different categories when comparing with the gold standard.

3.2.2 | Regression analysis

A logistic regression model (Table 5) was used to analyse the interaction of the different variables and showed a statistically significant lower probability of agreement on the grade for university faculty

TABLE 4 Absolute frequency and percentage of agreement comparing with gold standard

	Stage	p-Value	Grade	p-Value	Extent	p-Value
Overall	601 (68.7%)		717 (82.4%)		657 (75.5%)	
Current position		.341		.299		.144
University faculty	275 (71.4%)		311 (80.8%)		293 (76.1%)	
Specialist clinician	118 (65.6%)		155 (86.1%)		126 (70.0%)	
Postgraduate student	208 (68.2%)		251 (82.3%)		237 (77.7%)	
Experience (years)		.889		.710		.677
<5	277 (69.2%)		331 (82.8%)		304 (76.0%)	
≥5	312 (68.6%)		371 (81.5%)		340 (74.7%)	

Note: Comparison between categories by chi-square test.

TABLE 5 Logistic regression analysis

	Stage		Grade		Extent	
	OR (95% CI)	p-Value	OR (95% CI)	p-Value	OR (95% CI)	p-Value
Current position						
University faculty	1.35 (0.51–3.48)	.543	0.09 (0–0.49)	**.023***	2.67 (0.94–7.5)	.062
Postgraduate student	0.73 (0.28–1.8)	.450	0.12 (0.01–0.63)	**.045***	1.15 (0.41–3.02)	.777
Experience (years)						
≥5	1.15 (0.45–2.81)	.759	0.2 (0.01–1.07)	.127	0.85 (0.31–2.16)	.734

Note: Logistic regression model used. The odds ratio of the probability of agreement of university faculty and postgraduate students was calculated taking as a reference the clinicians.
Abbreviations: CI, coefficient interval; OR, odds ratio.
*Bold values are Statistically significant (*p* ≤ .05).

(OR = 0.09, 95% CI [0.00–0.49]; *p* = .023) and postgraduate residents (OR = 0.12, 95% CI [0.01–0.63]; *p* = .045) compared to specialists. In other words, the odds of reaching an agreement with the gold standard for grade was approximately 11 times lower for university faculty than clinicians and approximately 8 times lower in postgraduate students than clinicians.

4 | DISCUSSION

Every new classification system involves a learning curve, and this process may require some years. Hence, training, implementation, and practice are fundamental to avoid misclassification and incorrect treatment plans (Hefti & Preshaw, 2012). Furthermore, inconsistency in defining the different periodontitis categories can lead to incongruities in their prevalence, severity, and extent in epidemiological studies (Borrell & Papapanou, 2005; Page & Eke, 2007; Costa et al., 2009). For this reason, the purpose of this observational study was to assess the intra- and inter-examiner reliability in diagnosing periodontitis cases among specialist clinicians, faculty, and postgraduate students following the criteria of the 2018 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. The results from the present study have demonstrated a high level of validity and reliability in a sample of postgraduate students and specialists from EFP- and AAP-accredited programmes. These findings

seem to indicate that this new classification framework can be successfully used to diagnose periodontitis cases, which was reflected by the high concordance (>80%) between repeated diagnoses in the group of experts and by the high percentage of correct diagnoses in the general survey. However, it was more likely to reach a correct grade classification (85%), followed by the extent (75.5%), and, finally, the stage, which was the most difficult to assess (68.7%).

One of the objectives of this investigation was to determine whether expertise or the academic position could have an impact on the classification. However, the reported results show that neither the current position nor the experience of the periodontist influenced the outcomes, which further decreases the probability of diagnostic bias. We can interpret this finding in three different ways: first, this classification is simple enough to be accurately implemented outside of the academic setting; second, training does not require previous experience in the field; and third, transitioning from the previous classification can be done rather smoothly by the older generation of periodontists. Interestingly, the comparison between categories showed an effect of clinical experience on the assessment of the grade, as university faculty and postgraduate students demonstrated a slightly lower probability of determining the correct grade, compared to clinicians. Although specialist clinicians had a statistically significant greater probability of correct assessment of the grade than postgraduate students and university faculty, the odds were still very low and maybe not clinically relevant.

—WILEY—|

These results are in line with the study of Marini et al., (2020) where 30 participants of various education levels were recruited to evaluate 25 periodontitis cases. The sample consisted of undergraduate students, general dentists, and periodontal experts. Although the sample from the present study was larger, it did not include undergraduate students or general dentists. The periodontal experts' group in this study found a substantial agreement when comparing to a gold-standard classification for the stage (82%), the grade (72.4%), and the extent (84%). This result is comparable with the results from the present investigation, mainly for the grade (82.4%) and for the extent (75.5%), while it is slightly lower for staging (68.7%). Although both studies have shown good reliability for staging and grading, the lower percentages in the staging in this investigation can be explained by the presence of only stages III and IV, as the distinction between stage III versus IV seems to be more difficult than the distinction between stages I and II versus III and IV (Marini et al. 2020). Similar conclusions was made by Kornman and Papapanou (2020), who highlighted ground rules, clarifications, and "grey zones" for the clinical application of this new classification, emphasizing the need for a collective assessment of the potential complexity factors for the determination of the stage, rather than a mere "checking of a box" approach of isolated features. They also added that a correct implementation of the staging system requires a nuanced, thorough interpretation of a broad array of findings by a knowledgeable clinician. In this investigation, the participants were asked to use the algorithm developed by Tonetti and Sanz (2019) as an aid to reach the classification and to develop a treatment planning following the recently published treatment guidelines for the different stages of the disease (Sanz, Herrera, et al., 2020; Sanz, Papapanou, et al., 2020).

These "grey zones" were further highlighted in a similar online case-based study by Ravidà et al., in which 103 clinicians with prior training in the new periodontitis classification classified 10 severe periodontitis cases (Ravidà et al., 2021). The raters in this study achieved an inter-examiner agreement of 76% for stage, 82% for grade, and 84.8% for extent. This data are in line with the results from the present investigation as raters in both studies achieved very similar agreements for stage, grade, and extent. Moreover, the authors identified five common grey zone factors that reduced rater consistency by inviting the raters to submit queries concerning the selected cases. The said factors were the main determinants for identifying the stage, the definition of hopeless teeth, the differentiation between stage III and IV, the shift between the stages, and the assignment of the extent. In agreement to the authors' suggestions and in order to improve the classification agreement, identification of diagnostic challenges and complexities is required to promote the training of clinicians.

One of the main strengths of the present study is the sample, consisting of 174 participants from accredited periodontal training programme from both Europe and North America with different professional backgrounds, academic positions, and experiences. This study evaluated the inter-examiner agreement not only between participants but also in comparison to a gold-standard classification, determined by a group of seven internationally recognized experts,

three of which were directly involved in the development of this classification system. In addition, this study also assessed the consistency across time for every expert, which is also important in daily practice to establish a consistent treatment plan (Hefti & Preshaw, 2012).

The major limitation of this study was the use of an online survey sent indirectly by the programme director of each of the accredited programmes, which limited our ability to calculate the response rate, thus potentially entailing a selection bias. In fact, the survey was sent to 89 programmes and resulted in 174 participants completing the study. Although 174 replies seem low, sample size calculation had resulted in 90 participants, making the final sample adequate. However, we were able to estimate a response rate of 27% as 11 programmes confirmed their participation to the survey via email, resulting in 550 possible respondents. If we also consider that 5% of programmes did not confirm their participation but still forwarded the survey link to their alumni, this could result in 750 possible respondents. Moreover, the anonymity of the participants could have resulted in a selection bias across participants and specialist programmes, favouring those more familiar with the classification, thus overestimating the agreement. Furthermore, it could not be evaluated whether there was a clustering of responses by programme, as the survey was completely anonymous. Of the 11 programmes that confirmed their participation to the survey by email, there were six EFP and five AAP programmes. Furthermore, during the response acceptance phase of the survey (1 month), some important clarifications and updates were published, possibly changing the decision-making process between the first responders and the last responders (Sanz, Herrera, et al., 2020; Sanz, Papapanou, et al., 2020). In light of this new publication, one important clarification was that the assessment of extent should be made after stage determination and should describe the percentage of teeth at the stage-defining severity level (Sanz, Herrera, et al., 2020; Sanz, Papapanou, et al., 2020). This might explain the difference in the group of experts in assessing the stage and the grade between the first and the second time. Stage assessment may be particularly tricky, mostly for distinguishing between stages III and IV. The reason might be in calculating the number of teeth lost due to periodontitis, which constitutes the main severity factor differentiating stage III from stage IV. This assessment should also include hopeless teeth, which can be difficult to appropriately define (Sanz, Herrera, et al., 2020; Sanz, Papapanou, et al., 2020). Moreover, complexity factors such as masticatory dysfunction might be hard to diagnose. These factors can be considered rather judgemental in nature and could explain the difficulty in assessing the stage. Moreover, the periodontal chart used in this study reveals the probing pocket depth and the recession and indirectly permits the calculation of the CAL, which can lead to mathematical errors. Finally, the objective of this preliminary study was to assess the validity of the 2018 periodontitis classification in a specialist population. Another study including general dentists could be eventually contemplated in the future.

In conclusion, notwithstanding the low response rate, the potential selection bias, and clustering of the responses by programme, the data suggest that the use of the 2018 World Workshop on the

Classification of Periodontal and Peri-Implant Diseases and Conditions to properly assign the stage, grade, and extent of periodontitis demonstrates high inter-examiner reliability in experts in periodontology, specialized clinicians, and postgraduate students, regardless of their current position and experience. Given the limitations of this study, results should be interpreted with caution.

CONFLICT OF INTEREST

The authors declare no conflict of interests related to the content of this manuscript.

AUTHOR CONTRIBUTIONS

Lory Abrahamian, Andrés Pascual, Lucía Barallat, Cristina Valles, David Herrera, Mariano Sanz, José Nart, and Elena Figuero contributed to study conception and study design. Lory Abrahamian contributed to the survey preparation. Andrés Pascual-LaRocca, Lucía Barallat, Cristina Valles, David Herrera, Mariano Sanz, José Nart, and Elena Figuero contributed to the establishment of the gold-standard classification. Lory Abrahamian and Lucía Barallat contributed to data interpretation and manuscript drafting. Andrés Pascual, Lucía Barallat, Cristina Valles, David Herrera, Mariano Sanz, José Nart, and Elena Figuero contributed to critical revision of the manuscript. All authors approved the final version.

ETHICS STATEMENT

All patients were diagnosed and treated in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

DATA AVAILABILITY STATEMENT

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

ORCID

Lory Abrahamian https://orcid.org/0000-0003-0801-8523
Andrés Pascual-LaRocca https://orcid.org/0000-0002-1223-7764
Lucía Barallat https://orcid.org/0000-0002-9250-5625
Cristina Valles https://orcid.org/0000-0002-9359-1705
David Herrera https://orcid.org/0000-0002-2258-6597
Mariano Sanz https://orcid.org/0000-0002-6293-5755
José Nart https://orcid.org/0000-0002-2363-4992
Elena Figuero https://orcid.org/0000-0002-3129-1416

REFERENCES

Armitage, G. C. (1999). Development of a classification system for periodontal diseases and conditions. *Annals of Periodontology*, 4(1), 1–6. https://doi.org/10.1902/annals.1999.4.1.1

Armitage, G. C. (2014). AAP centennial commentary: Theme 3: Evolution and application of classification systems for periodontal diseases–A retrospective commentary. *Journal of Periodontology*, 85(3), 369–371. https://doi.org/10.1902/jop.2014.130700

Borrell, L. N., & Papapanou, P. N. (2005). Analytical epidemiology of periodontitis. *Journal of Clinical Periodontology*, 32(Suppl. 6), 132–158. https://doi.org/10.1111/j.1600-051X.2005.00799.x

Caton, J. G., Armitage, G., Berglundh, T., Chapple, I. L. C., Jepsen, S., Kornman, K. S., Mealey, B. L., Papapanou, P. N., Sanz, M., & Tonetti, M. S. (2018). A new classification scheme for periodontal and peri-implant diseases and conditions—Introduction and key changes from the 1999 classification. *Journal of Clinical Periodontology*, 45 Suppl. 20, S1–S8. https://doi.org/10.1111/jcpe.12935

Cohen, J. F., Korevaar, D. A., Altman, D. G., Bruns, D. E., Gatsonis, C. A., Hooft, L., Irwig, L., Levine, D., Reitsma, J. B., de Vet, H. C. W., & Bossuyt, P. M. M. (2016). STARD guidelines for reporting diagnostic accuracy studies: Explanation and elaboration. *BMJ Open*, 6(11), 1–17. https://doi.org/10.1136/bmjopen-2016-012799

Costa, F. O., Guimarães, A. N., Cota, L. O., Pataro, A. L., Segundo, T. K., Cortelli, S. C., & Costa, J. E. (2009). Impact of different periodontitis case definitions on periodontal research. *Journal of Oral Science*, 51, 199–206. https://doi.org/10.2334/josnuns.51.199

Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. (2nd ed, pp. 38–46). New York, NY: John Wiley.

Graetz, C., Mann, L., Krois, J., Sälzer, S., Kahl, M., Springer, C., & Schwendicke, F. (2019). Comparison of periodontitis patients' classification in the 2018 versus 1999 classification. *Journal of Clinical Periodontology*, 46, 908–917. https://doi.org/10.1111/jcpe.13157

Hamp, S. E., Nyman, S., & Lindhe, J. (1975). Periodontal treatment of multirooted teeth. Results after 5 years. *Journal of Clinical Periodontology*, 2, 126–135.

Hefti, A. F., & Preshaw, P. M. (2012). Examiner alignment and assessment in clinical periodontal research. *Periodontology 2000*, 59(1), 41–60. https://doi.org/10.1111/j.1600-0757.2011.00436.x

Kornman, K. S., & Papapanou, P. N. (2020). Clinical application of the new classification of periodontal diseases: Ground rules, clarifications and "gray zones". *Journal of Periodontology*, 91(3), 352–360. https://doi.org/10.1002/JPER.19-0557

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.

Marini, L., Tonetti, M. S., Nibali, L., Rojas, M. A., Aimetti, M., Cairo, F., Cavalcanti, R., Crea, A., Ferrarotti, F., Graziani, F., Landi, L., Sforza, N. M., Tomasi, C., & Pilloni, A. (2020). The staging and grading system in defining periodontitis cases: Consistency and accuracy among periodontal experts, general dentists and undergraduate students. *Journal of Clinical Periodontology*, 48, 205–215. https://doi.org/10.1111/jcpe.13406

Miller, P. D. (1985). A classification of marginal tissue recession. *International Journal of Periodontics and Restorative Dentistry*, 5, 9–13.

Milward, M. R., & Chapple, I. L. C. (2003). Classification of periodontal diseases: Where were we? Where are we now? Where are we going? *Dental Update*, 30(1), 37–44. https://doi.org/10.12968/denu.2003.30.1.37

Orban, B. (1942). Classification and nomenclature of periodontal diseases (1) (based on pathology, etiology, and clinical picture). *The Journal of Periodontology*, 13(2), 88–91. https://doi.org/10.1902/jop.1942.13.2.88

Page, R. C., & Eke, P. I. (2007). Case definitions for use in population-based surveillance of periodontitis. *Journal of Periodontology*, 78(1), 1387–1399. https://doi.org/10.1902/jop.2007.060264

Papapanou, P. N., Sanz, M., Buduneli, N., Dietrich, T., Feres, M., Fine, D. H., Flemmig, T. F., Garcia, R., Giannobile, W. V., Graziani, F., Greenwell, H., Herrera, D., Kao, R. T., Kebschull, M., Kinane, D. F., Kirkwood, K. L., Kocher, T., Kornman, K. S., Kumar, P. S., … Tonetti, M. S. (2018). Periodontitis: Consensus report of workgroup 2 of the 2017 World Workshop on the Classification of Periodontal

and Peri-Implant Diseases and Conditions. *Journal of Clinical Periodontology*, *45 Suppl. 20*, S162–S170. https://doi.org/10.1111/jcpe.12946

Ravidà, A., Travan, S., Saleh, M., Greenwell, H., Papapanou, P. N., Sanz, M., Tonetti, M., Wang, H. L., & Kornman, K. (2021). Agreement among international periodontal experts using the 2017 World Workshop classification of periodontitis. *Journal of Periodontology*, *92*(12), 1675–1686. https://doi.org/10.1002/JPER.20-0825

Sanz, M., Herrera, D., Kebschull, M., Chapple, I., Jepsen, S., Berglundh, T., Sculean, A., Tonetti, M. S., Merete Aass, A., Aimetti, M., Kuru, B., Belibesakis, G., Blanco, J., Bol-van den Hil, E., Bostanci, N., Bozic, D., Bouchard, P., Budunelli, N., … Wennström, J. (2020). Treatment of stage I-III periodontitis – The EFP S3 level clinical practice guideline. *Journal of Periodontology*, *47*, 4–60. https://doi.org/10.1111/jcpe.13290

Sanz, M., Papapanou, P. N., Tonetti, M. S., Greenwell, H., & Kornman, K. (2020). Guest editorial: Clarifications on the use of the new classification of periodontitis. *Journal of Clinical Periodontology*, *47*(6), 658–659. https://doi.org/10.1111/jcpe.13286

Tonetti, M. S., Greenwell, H., & Kornmann, K. S. (2018). Staging and grading of periodontitis: Framework and proposal of a new classification and case definition. *Journal of Clinical Periodontology*, *44 Suppl. 20*, S149–S161. https://doi.org/10.1111/jcpe.12945

Tonetti, M. S., & Sanz, M. (2019). Implementation of the new classification of periodontal diseases: Decision-making algorithms for clinical practice and education. *Journal of Clinical Periodontology*, *46*(4), 398–405. https://doi.org/10.1111/jcpe.13104

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.