Check for updates

Received: 27 February 2023 | Revised: 24 April 2023 | Accepted: 9 May 2023

DOI: 10.1002/JPER.23-0129

JOURNAL OF Periodontology

ORIGINAL ARTICLE

Reliability assessment of the 2018 classification case definitions of peri-implant health, peri-implant mucositis, and peri-implantitis

Lorenzo Marini[1] | **Cristiano Tomasi**[2] | **Rodolfo Gianserra**[3] | **Filippo Graziani**[4] | **Luca Landi**[5] | **Mauro Merli**[6,7] | **Luigi Nibali**[8] | **Mario Roccuzzo**[9,10,11] | **Nicola M. Sforza**[12] | **Maurizio S. Tonetti**[13] | **Federico Deli**[1] | **Piero Papi**[1] | **Bianca Di Murro**[1] | **Mariana A. Rojas**[1] | **Andrea Pilloni**[1]

[1]Section of Periodontics, Department of Oral and Maxillofacial Sciences, Sapienza University of Rome, Rome, Italy

[2]Department of Periodontology, Institute of Odontology, The Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden

[3]Private Practice, Rome and Campobasso, Italy

[4]Sub-Unit of Periodontology, Halitosis and Periodontal Medicine, Department of Surgical, Medical and Molecular Pathology and Critical Care Medicine, University of Pisa, Pisa, Italy

[5]Private Practice, Rome and Verona, Italy

[6]Politecnico delle Marche, Ancona, Italy

[7]Private Practice, Rimini, Italy

[8]Periodontology Unit, Centre for Host-Microbiome Interactions, Faculty of Dental and Craniofacial Sciences, King's College London, London, UK

[9]Division of Maxillofacial Surgery, University of Torino, Torino, Italy

[10]Department of Periodontics and Oral Medicine, University of Michigan, Ann Arbor, Michigan, USA

[11]Private Practice, Torino, Italy

[12]Private Practice, Bologna, Italy

[13]Shanghai PerioImplant Innovation Center, Department of Oral and Maxillo-facial Implantology, Shanghai Key Laboratory of Stomatology, National Clinical Research Centre for Oral Diseases, Shanghai Ninth People Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Correspondence
Lorenzo Marini, Section of Periodontics, Department of Oral and Maxillofacial Sciences, Sapienza University of Rome, #6 Via Caserta, Rome 00161, Italy.
Email: lorenzo.marini@uniroma1.it

Abstract

Background: The purpose of this study was to evaluate the reliability and accuracy in the assignment of the case definitions of peri-implant health and diseases according to the 2018 Classification of Periodontal and Peri-implant Diseases and Conditions.

Methods: Ten undergraduate students, 10 general dentists, and 10 experts in implant dentistry participated in this study. All examiners were provided with clinical and radiographic documentation of 25 dental implants. Eleven out the 25 cases were also accompanied by baseline readings. Examiners were asked to define all cases using the 2018 classification case definitions. Reliability among examiners was evaluated using the Fleiss kappa statistic. Accuracy was estimated

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2023 The Authors. *Journal of Periodontology* published by Wiley Periodicals LLC on behalf of American Academy of Periodontology.

using percentage of complete agreement and quadratic weighted kappa for pairwise comparisons between each rater and a gold standard diagnosis.

Results: The Fleiss kappa was 0.50 (95% CI: 0.48 to 0.51) and the mean quadratic weighted kappa value was 0.544. Complete agreement with the gold standard diagnosis was achieved in 59.8% of the cases. Expertise in implantology affected accuracy positively ($p < 0.001$) while the absence of baseline readings affected it negatively ($p < 0.001$).

Conclusion: Both reliability and accuracy in assigning case definitions to dental implants according to the 2018 classification were mostly moderate. Some difficulties arose in the presence of specific challenging scenarios.

KEYWORDS

classification, dental implants, diagnosis, disease, health, peri-implantitis, reproducibility of results

1 | INTRODUCTION

Patients who received implant-supported restorative therapy may experience biological complications represented by inflammatory conditions of the soft and hard peri-implant tissues, induced by the bacterial biofilm.[1,2] The 2018 Classification of Periodontal and Peri-implant Diseases and Conditions proposed a new definition of these pathologies, aiming at introducing a uniform classification for peri-implant health, peri-implant mucositis, and peri-implantitis.[3,4] Indeed, the lack of consensus on the diagnosis of peri-implant health and diseases resulted in a huge heterogeneity in the reported prevalence rate of these disorders and led to misdiagnosis and over- and under-treatment of the disease.[5] Moreover, peri-implant tissue health status has been recently defined as an essential outcome domain that should be captured in all implant clinical trials, and the 2018 classification case definitions have been used in the European Federation of Periodontology guidelines to drive prevention and treatment of peri-implant diseases.[6,7]

According to the 2018 classification, the distinction between peri-implant health and peri-implant mucositis is based on the presence or absence of the following: (1) clinical signs of inflammation (e.g., gingival erythema, edema, and changes in soft tissue consistency); (2) bleeding (BoP) and/or suppuration (SoP) on gentle probing; and (3) increased probing pocket depth (PD). Peri-implantitis—in addition to the previous clinical signs of inflammation of the mucosa—is characterized by radiographic evidence of progressive bone loss.[3,4] A secondary case definition was proposed, in the absence of longitudinal data, that includes bleeding on probing and/or suppuration on probing at ≥1 site, probing depth ≥6 mm, and bone level ≥3 mm.[3,4]

As with any classification system, it is critical to estimate external validity of the 2018 classification case definitions when applied in research and clinical practice in order to prove they are exhaustive, non-ambiguous, simple, and have high reproducibility.[8] Recently, the reproducibility of the 2018 classification case definitions of periodontitis and of gingival recession defects and gingival phenotype has been assessed.[9–12] However, no previous study has assessed the reproducibility of the new case definitions of peri-implant health and diseases. Therefore, the purpose of this survey was to evaluate the agreement between examiners with different levels of education and expertise in assigning the 2018 classification case definitions to dental implants and their accuracy against a reference diagnosis.

2 | MATERIALS AND METHODS

2.1 | Study design

This investigation was designed to test the reliability of the 2018 classification case definitions of peri-implant health, peri-implant mucositis, and peri-implantitis among three groups of raters divided according to their level of education and training in implantology.[3,4] Additionally, the examiners' case definitions were compared to the reference diagnosis assigned by a gold standard examination to estimate accuracy. Photographs, clinical, and radiographic data of 25 implant cases were used for this study. The study was prepared following the Guidelines for Reporting Reliability and Agreement Studies (GRRAS).[13]

2.2 | Ethical considerations

Anonymous and non-identifiable data collected from subjects who received implant-supported restorative therapy in the context of routine care at the Section of Periodontology of Sapienza University of Rome were used in

19433670, 0, Downloaded from https://aap.onlinelibrary.wiley.com/doi/10.1002/JPER.23-0129 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

this study. Patients provided their informed consent. The study protocol was approved by the Ethical Committee of Sapienza University of Rome (#6973; 944/2022; approval date: 07/12/2022), in accordance with the Declaration of Helsinki of 1975, revised in 2013.

2.3 | Examiners

A total of 30 participants from different levels of education and training in implant dentistry were selected as follows:

(a) Ten final year undergraduate dental students from Sapienza University of Rome. They attended courses in periodontology and implantology during their university career.
(b) Ten general dentists, with at least 10 years of clinical experience in general dentistry, who had graduated from Sapienza University of Rome. They were clinicians whose practice is not limited to a specific discipline and who are not certified by a recognized specialty board.
(c) Ten experts in implant dentistry (R.G., F.G., L.L., M.M., L.N., M.R., N.M.S., M.S.T., F.D., A.P.) who are active members of or certified by the Italian Society of Periodontology and Implantology, whose expertise has been recognized by peers.

The first group was recruited randomly from the student roster, while the other two groups were selected at the direct invitation of a study team member (L.M).

All examiners were informed about the purpose of the study and their participation was voluntary.

2.4 | Gold standard diagnosis

Two different examiners were designated (C.T. and Jan Derks), from among the participants in Working Group 4 on the Classification of Peri-implant Diseases and Conditions in the context of the 2017 World Workshop, to assign the "true" diagnosis to all cases.[3] They independently evaluated the 25 cases, and cases that were not uniformly defined were discussed until a unanimous decision was reached.

2.5 | Clinical cases

The examiners were provided with a PDF document containing 25 numbered slides. Each slide showed one dental implant case, providing the following data needed for case definition:

- time (years) from the delivery of the implant-supported prosthetic reconstruction;
- probing depth (mm) measured at six sites per implant with a manual periodontal probe* without removal of the reconstruction;
- bleeding on probing (BoP) (no/yes) recorded at six sites per implant within 15 s following probing;
- suppuration on probing (SoP) (no/yes) assessed at six sites per implant within 15 s following pocket probing;
- two intraoral photographs (one buccal and the other palatal/lingual) showing the clinical aspect of the dental implant and the soft peri-implant tissues;
- a long cone, parallel technique, periapical radiograph of the dental implant. To allow assessments of the bone level, each radiograph was provided with a millimeter ruler whose beginning was at the level of the most coronal point of the intraosseous part of the implant. The implant length was used for the ruler calibration.

Eleven out of the 25 cases were also accompanied by baseline readings obtained from patient files, consisting of the following:

- probing depth (mm) measured at six sites per implant with a manual periodontal probe* at the time of superstructure placement;
- a long cone, parallel technique, periapical radiograph of the dental implant taken at 1 year after delivery of the prosthetic restoration.

A representative example of case documentation is shown in Figure 1.

All clinical measurements were performed by a single calibrated investigator (L.M.) who was not involved in the assessment of reliability. Similarly, intraoral photographs and radiographs were taken by clinicians other than examiners.

The document including all 25 clinical cases is provided as supporting information (Supplementary Material S1 in the online *Journal of Periodontology*).

2.6 | Assignment of case definitions

Prior to the distribution of the PDF document containing the 25 cases for evaluation, the examiners were provided with detailed information on the study procedures. Moreover, they were provided with two cases, not included in the study, for explaining the case documentation and evaluation modalities. In addition, they received a predesigned data collection sheet to record their diagnoses.

* PCP15 (Hu-Friedy, Chicago, IL, USA).

19433670, 0, Downloaded from https://aap.onlinelibrary.wiley.com/doi/10.1002/JPER.23-0129 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

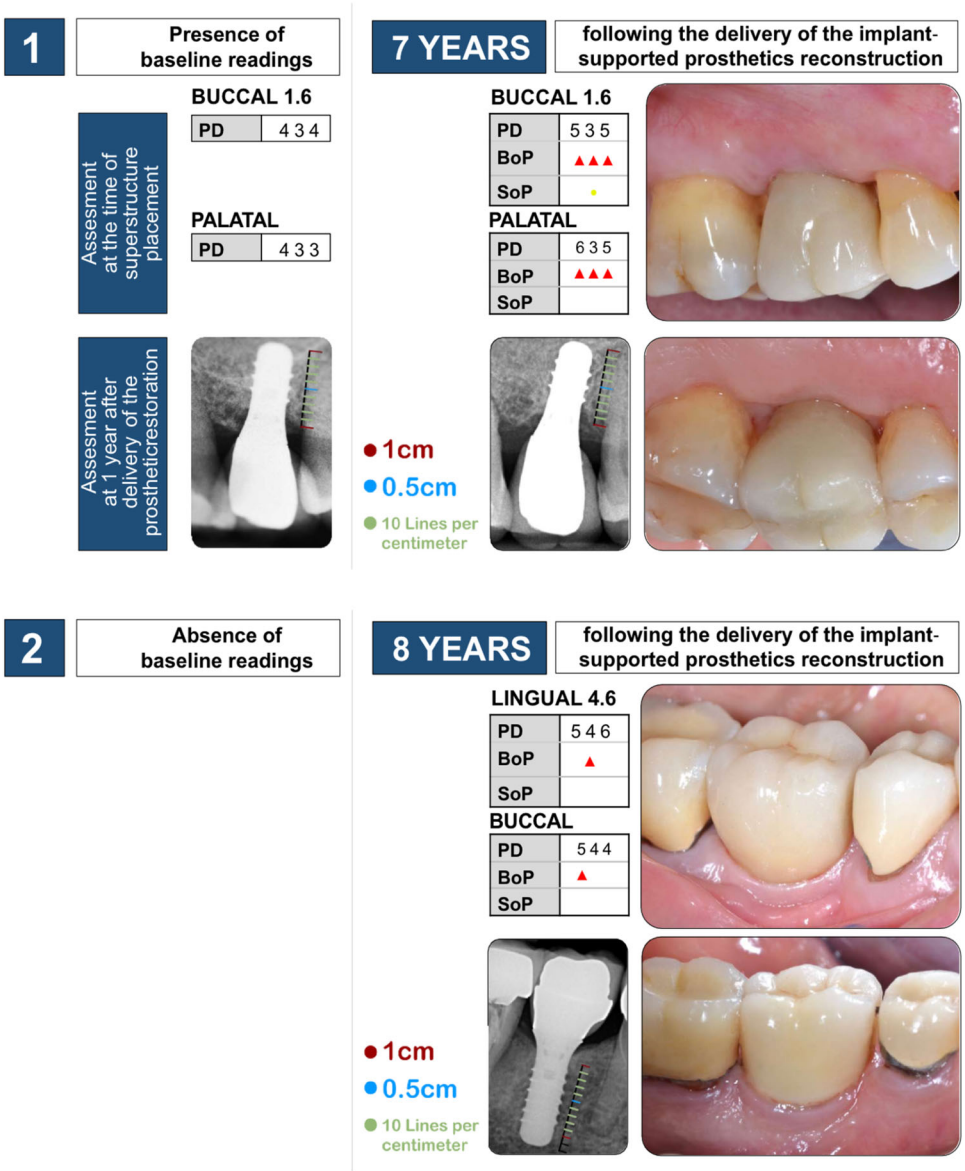

FIGURE 1 Representative examples of documentation provided for each dental implant in presence (case 1) or in absence (case 2) of baseline readings. PD, probing depth; BoP, bleeding on probing; SoP, suppuration on probing

None of the 30 participants were aware of the cases prior to the evaluation, nor did they receive any other information or guidance during the assessment.

Examiners assigned case definitions to dental implants through direct (when longitudinal data were available) and/or indirect evidence according to the 2018 classification.[3] They accomplished their task independently and blindly to each other, from their own workstations and without time limitations.

Additional training and calibration on 2018 classification case definitions of peri-implant health and diseases were intentionally not provided to examiners prior to the study. However, during the assessments, all participants were allowed to access a summary of the parameters for case definition of each peri-implant health status, prepared by a study team member (L.M). A summary is provided as Supplementary Material S2 in the online *Journal of Periodontology*.

2.7 | Outcomes

Accuracy, defined as the reliability between the case definitions provided by each examiner and those assigned by the gold standard diagnosis, was considered as the primary outcome. The secondary outcomes, considered as potential explanatory outcomes, were (1) the reliability among examiners (overall and by group) in defining peri-implant health status; and (2) agreement with the gold standard case definition in relation to the presence or

19433670, 0, Downloaded from https://aap.onlinelibrary.wiley.com/doi/10.1002/JPER.23-0129 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

absence of baseline readings and the education and clinical experience of the examiners.

2.8 | Data analysis

Continuous variables were described by means (± standard deviation) and categorical variables by frequency distributions (percentage).

The reliability between each examiner and the gold standard was estimated by quadratic weighted kappa.[14] The inter-examiner reliability was evaluated using the Fleiss kappa statistic.[15]

The kappa values have been interpreted as follows: poor agreement = < 0.00; slight agreement = 0.00 to 0.20; fair agreement = 0.21 to 0.40; moderate agreement = 0.41to 0.60; substantial agreement = 0.61to 0.80; and almost perfect agreement = 0.81 to 1.00.[16]

Statistically significant differences between expected and observed frequencies in complete agreement with the reference diagnosis according to the examiner group and the presence or absence of baseline readings were assessed using the chi-squared test. The significance level (α) was set at 0.05.

All analyses were performed using dedicated software[†].

2.9 | Sample size

The numbers of clinical cases required for kappa statistics for two observers (each examiner vs. the gold standard diagnosis) and three categories (peri-implant health, peri-implant mucositis, and peri-implantitis) were estimated using the confidence interval perspective, using the CI3Cats function of the kappaSize package for the R environment for statistical computing[‡].[17] The anticipated value of kappa was set at 0.50, the lower bound of the 95% CI was set at 0.20, and the upper bound at 0.80. In addition, the anticipated prevalence of peri-implant health, peri-implant mucositis, and peri-implantitis were set at 0.30, 0.40, and 0.30, respectively. Using the above-mentioned parameters, a minimum sample of 25 subjects was necessary.

Since in reliability studies the number of clinical cases has a much greater impact on consistency than the number of examiners,[18] the sample of examiners was based on generalizability and feasibility. Hence, according to comparable studies,[9,19,20] 30 evaluators (10 per group) were included in this investigation.

[†] IBM SPSS Statistics (IBM SPSS Statistics for Macintosh, Version 25.0. Armonk, NY: IBM Corp).

[‡] R: R Foundation for Statistical Computing, Vienna, Austria.

3 | RESULTS

3.1 | Descriptive characteristics of implant cases

Twenty-five dental implants with a mean of 9.2 ± 5.9 years from delivery of the implant-supported prosthetic reconstruction were examined in the present study. The sample consisted of 15 (60%) implants in the maxilla and 10 (40%) in the mandible. Twenty-three (92%) implants were in the posterior area, while two (8%) were in the anterior area (canine-canine). The retention of suprastructure was screw-retained in seven (28%) cases and cemented in 18 (72%) cases, while the design of the suprastructure was single unit or multi-unit in 19 (76%) and six (24%) cases, respectively. Eleven (44%) implants were provided with baseline clinical and radiographic data.

According to the diagnosis made by the gold standard examination, three (12%) implants were healthy, while 16 (64%) cases were affected by peri-implant mucositis and six (24%) cases by peri-implantitis. Presence of bone loss (≥0.5 mm) was identified in 18 (72%) cases.

3.2 | Agreement with gold standard case definitions

Table 1 provides the case definitions assigned by the gold standard examination and the rationale for each diagnosis.

Figure 2 shows the relative proportions of peri-implant health, peri-implant mucositis, and peri-implantitis assigned by examiners for every case along with the case definitions provided by the gold standard diagnosis.

Values of quadratic weighted kappa for pairwise comparisons of each examiner against the gold standard are presented in Table 2. The mean value of the quadratic weighted kappa (k = 0.544) was interpreted as moderate agreement. The frequencies and percentages of poor, slight, fair, moderate, substantial, and almost perfect agreements achieved by pairwise comparisons are shown in Figure 3(A). The experts in implant dentistry were more accurate than general dentists and undergraduate dental students, with a higher percentage of examiners showing substantial (40% vs. 20% and 10%, respectively) and almost perfect (10% vs. 0% and 0%, respectively) agreement with the gold standard.

Frequencies and percentages of complete agreement with gold standard case definitions are shown in Figure 3(B). Complete agreement with the gold standard diagnosis was achieved in 59.8% of the cases. There was a statistically significant difference in the ability to assign an accurate case definition based on examiner education and

19433670, 0, Downloaded from https://aap.onlinelibrary.wiley.com/doi/10.1002/JPER.23-0129 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Case definitions assigned by the gold standard examination and the rationale for each diagnosis.

	Case 1[a]	Case 2[b]	Case 3[c]	Case 4[d]	Case 5[e]	Case 6[f]	Case 7[g]	Case 8[h]	Case 9[i]	Case 10[j]	Case 11[k]	Case 12[l]	Case 13[m]	Case 14[n]	Case 15[o]	Case 16[p]	Case 17[q]	Case 18[r]	Case 19[s]	Case 20[t]	Case 21[u]	Case 22[v]	Case 23[w]	Case 24[x]	Case 25[y]
Gold standard definition	P	M	H	M	H	M	P	M	M	M	M	M	P	P	M	M	P	M	P	M	M	M	M	H	M
Justification																									
Visible signs of inflammation (red, swollen, soft consistency)	N	N	N	N	N	Y	N	N	N	Y	Y	N	Y	N	Y	N	Y	N	Y	Y	Y	Y	N	N	Y
Bleeding on probing	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y
Suppuration on probing	Y	N	N	Y	N	Y	N	N	N	N	N	N	Y	N	N	N	Y	N	Y	Y	Y	Y	N	N	Y
Increasing probing pocket depths as compared to measurements obtained at placement of the suprastructure	Y	NA	Y	NA	N	NA	Y	NA	Y	NA	Y	N	NA	Y	NA	N	NA	Y	NA	Y	NA	NA	NA	NA	NA
Bone loss in relation to the radiographic bone level assessment at 1 year following the delivery of the implant-supported prosthetics reconstruction	N	NA	Y	NA	N	NA	Y	NA	Y	NA	N	N	NA	Y	NA	N	NA	N	NA	N	NA	NA	NA	NA	NA

(Continues)

19433670, 0, Downloaded from https://aap.onlinelibrary.wiley.com/doi/10.1002/JPER.23-0129 by Tesi, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

	Case 1[a]	Case 2[b]	Case 3[c]	Case 4[d]	Case 5[e]	Case 6[f]	Case 7[g]	Case 8[h]	Case 9[i]	Case 10[j]	Case 11[k]	Case 12[l]	Case 13[m]	Case 14[n]	Case 15[o]	Case 16[p]	Case 17[q]	Case 18[r]	Case 19[s]	Case 20[t]	Case 21[u]	Case 22[v]	Case 23[w]	Case 24[x]	Case 25[y]
Probing depths ≥6 mm in conjunction with profuse bleeding	NA	N	NA	Y	NA	Y	NA	N	NA	Y	NA	NA	N	NA	Y	NA	Y	NA	Y	NA	Y	Y	N	N	N
Radiographic evidence of bone level ≥3 mm	NA	N	NA	N	NA	N	NA	N	NA	N	NA	NA	N	NA	N	NA	Y	NA	Y	NA	N	Y	N	N	N
Radiographic bone level in mm	2	0	1	2	0	2	1	0	1	2	0	1	1	3	1	0	3	1	3	0	1	2	1	0	1

[a]No signs of inflammation; BoP+; SoP+; PD increase; Bone loss
[b]BoP+; PD ≥6 mm; Bone level; < 3 mm
[c]No signs of inflammation; PD change in one site; Bone loss; 1 mm
[d]No signs of inflammation; Bop+; SoP+; PD ≥6 mm; Bone level; < 3 mm
[e]No signs of inflammation; No PD increase; No bone loss
[f]Signs of inflammation; Bop+; SoP+; PD ≥6 mm; Bone level; < 3 mm
[g]No signs of inflammation; BoP+; PD increase; Bone loss 1 mm
[h]No signs of inflammation; BoP+; PD < 6 mm; No bone loss
[i]No signs of inflammation; BoP+; 1 mm change in PD at 1 site; Bone loss 1 mm (but projection prescription different)
[j]Signs of inflammation; BoP+; SoP+; PD ≥6 mm; Bone level; < 3 mm
[k]Signs of inflammation; BoP+; PD increase; No bone loss
[l]No signs of inflammation; BoP+; PD < 6 mm; No bone loss
[m]Signs of inflammation; BoP+; SoP+; PD < 6; Bone level; < 3 mm
[n]No signs of inflammation; BoP+; PD increase; Bone loss; ≥3 mm
[o]Signs of inflammation; Bop+; PD ≥6 mm; Bone level; < 3 mm
[p]No signs of inflammation; BoP+; No PD increase; -No bone loss
[q]Signs of inflammation; Bop+; SoP+; PD ≥6 mm; Bone level; ≥3 mm
[r]No signs of inflammation; BoP+ half sites (mesial); PD increase; -No clear bone loss (baseline radiograph over-projected)
[s]Signs of inflammation; Bop+; SoP+; PD ≥6 mm; Bone level; ≥3 mm
[t]Signs of inflammation; Bop+; SoP+; PD increase in 1 site; No bone loss
[u]Signs of inflammation; Bop+; SoP+; PD ≥6 mm; Bone level; < 3 mm
[v]Signs of inflammation; Bop+; SoP+; PD ≥6 mm; Bone level; < 3 mm
[w]No signs of inflammation; BoP+; PD < 6 mm; No bone loss
[x]No signs of inflammation; PD < 6 mm; -No bone loss
[y]Signs of inflammation; Bop+; SoP+; PD ≥6 mm; Bone level < 3 mm

Abbreviations: BoP, bleeding on probing; H, peri-implant health; M, mucositis; N, no; NA, not available; P, peri-implantitis; PD, probing depth; SoP, suppuration on probing; Y, yes.

19433670, 0, Downloaded from https://aap.onlinelibrary.wiley.com/doi/10.1002/JPER.23-0129 by Tesa, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

FIGURE 2 Relative proportions of peri-implant health, peri-implant mucositis, and peri-implantitis assigned by examiners for every case along with the case definitions provided by the gold standard examination

clinical experience. Indeed, implant experts performed better than general dentists (69.2% vs. 58.0%, $p = 0.009$) and undergraduate students (69.2% vs. 52.4%, $p < 0.001$). Conversely, there was no significant difference between general dentists and undergraduate dental students ($p = 0.208$). Moreover, presence or absence of baseline readings statistically significantly affected the possibility of complete agreement with the gold standard diagnosis (presence vs. absence of baseline readings, 65.1% vs. 55.7%; $p < 0.001$) (Figure 3C).

3.3 | Inter-examiner agreement

Table 3 presents the results of the Fleiss kappa statistic relating to the agreement between the overall group of

19433670, 0, Downloaded from https://aap.onlinelibrary.wiley.com/doi/10.1002/JPER.23-0129 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

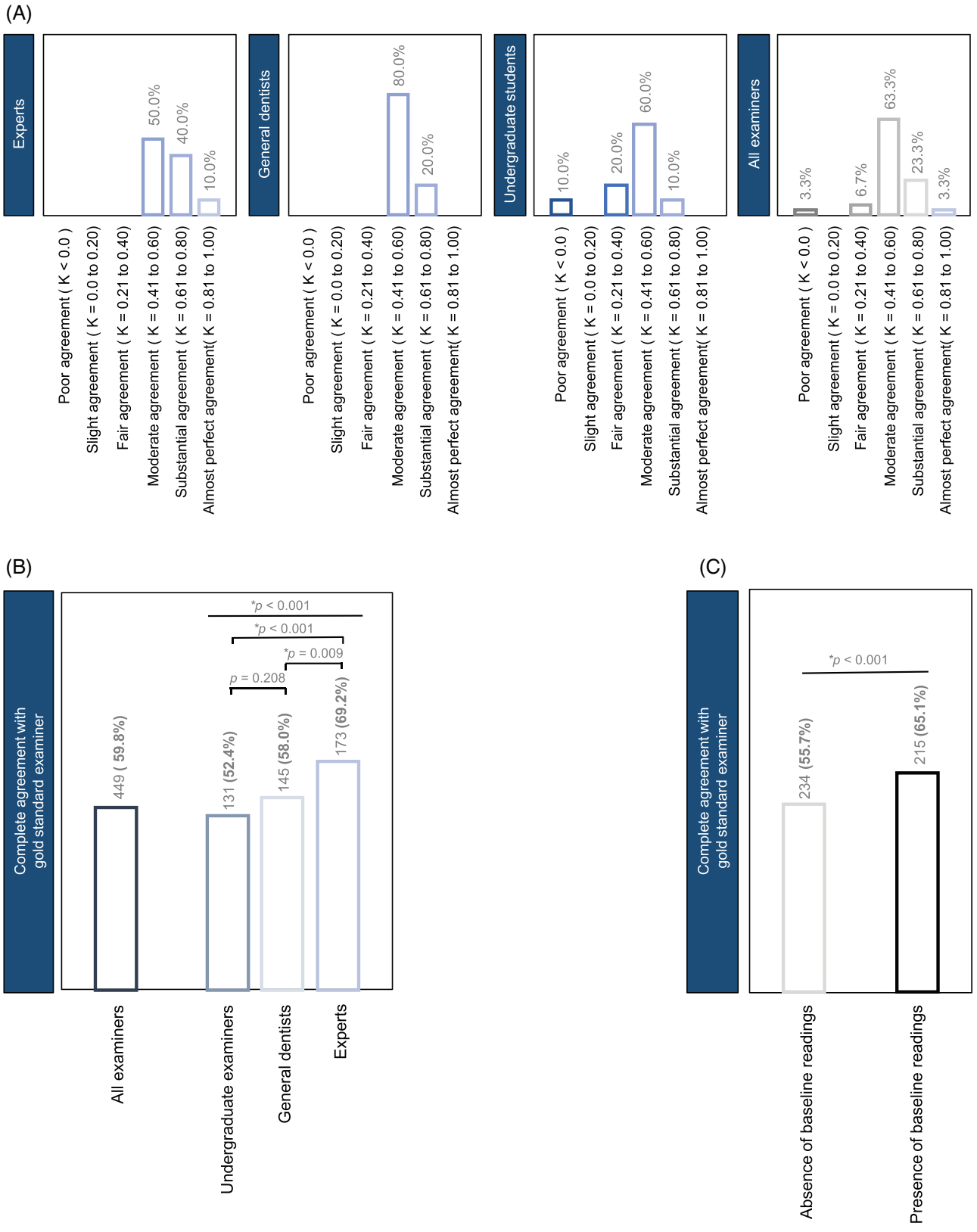

FIGURE 3 (A) Percentages of poor, slight, fair, moderate, substantial, and almost perfect agreement achieved by pairwise comparisons (gold standard vs. each examiner), for groups of examiners stratified by education and clinical experience (experts, general dentists, undergraduate students) and for all examiners. (B,C) Frequencies and percentages of case definitions consistent with those of the gold standard diagnosis, with significance assessed by the chi-squared test, according to the education and clinical experience of examiners (B) and the presence or absence of baseline readings (C). *$p < 0.05$

19433670, 0, Downloaded from https://aap.onlinelibrary.wiley.com/doi/10.1002/JPER.23-0129 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Quadratic weighted kappa for pairwise comparisons of each examiner against gold standard examiner (95% CIs).

Group	Examiner number	Quadratic weighted kappa	95% CIs (lower, upper bound)
Undergraduate students	1	0.395	0.143, 0.646
	2	0.454	0.225, 0.683
	3	0.333	−0.010, 0.676
	4	0.600	0.355, 0.846
	5	0.455	0.190, 0.719
	6	0.566	0.307, 0.825
	7	0.417	0.163, 0.671
	8	−0.007	−0.426, 0.412
	9	0.510	0.272, 0.748
	10	0.620	0.359, 0.880
General dentists	1	0.472	0.246, 0.697
	2	0.560	0.290, 0.830
	3	0.590	0.343, 0.837
	4	0.488	0.240, 0.736
	5	0.516	0.285, 0.747
	6	0.606	0.380, 0.831
	7	0.617	0.393, 0.841
	8	0.500	0.254, 0.746
	9	0.513	0.283, 0.743
	10	0.654	0.373, 0.936
Experts	1	0.583	0.305, 0.861
	2	0.558	0.325, 0.792
	3	0.703	0.449, 0.957
	4	0.532	0.296, 0.767
	5	0.659	0.415, 0.903
	6	0.581	0.342, 0.820
	7	0.507	0.269, 0.744
	8	0.727	0.521, 0.933
	9	0.675	0.435, 0.915
	10	0.941	0.826, 1.000

TABLE 3 Fleiss kappa statistics indicating the agreement between the overall group of examiners and between each group of observers.

Examiners	Kappa	95% CIs (lower, upper bound)
All	0.50	0.48, 0.51
Groups		
Experts ($n = 10$)	0.52	0.48, 0.56
General dentists ($n = 10$)	0.53	0.48, 0.57
Undergraduate students ($n = 10$)	0.46	0.41, 0.50

examiners and between each group of observers (undergraduate students, general dentists, and experts).

Inter-rater agreement was simply moderate (k = 0.50 [0.48, 0.51]), with similar results within the three groups (undergraduate students: k = 0.46 [0.41, 0.50]; general dentists: k = 0.53 [0.48, 0.57]; and implant experts: k = 0.52 [0.48, 0.56]).

The kappa value was higher when cases were diagnosed by the gold standard examination as peri-implant health (k = 0.81 [0.70, 0.91]) than when they were classified as peri-implant mucositis (k = 0.43 [0.31, 0.56]) and peri-implantitis (k = 0.56 [0.25, 0.88]).

Furthermore, agreement among all examiners in the presence or absence of baseline readings resulted in k values of 0.48 (0.43, 0.52) and 0.47 (0.43, 0.50), respectively.

4 | DISCUSSION

4.1 | Main findings

Classifications should allow accurate and reproducible definitions, as inconsistent assessment can generate serious consequences, including misdiagnosis and inappropriate treatment.[21,22] Therefore, this study was of great interest, as it aimed at assessing for the first time the reliability and the accuracy in assigning the case definitions of peri-implant health, peri-implant mucositis, and peri-implantitis according to the 2018 classification. The main findings of this study were the following: (1) the examiners were accurate in just over half of the cases, mostly showing moderate agreement with the gold standard diagnosis; (2) accuracy was affected both by the presence/absence of longitudinal data and by examiner training and clinical experience; and (3) agreement among examiners was moderate, with comparable results in each group.

With respect to the primary outcome, pairwise comparisons between each examiner and the gold standard showed a mean quadratic weighted kappa value of 0.544. This was interpreted as moderate agreement and it was close to the expected value. Also with regard to the accuracy of the diagnosis, the percentage of complete agreement with the case definitions provided by the gold standard examiners was 59.8%. Taken together, these data seem to highlight the need for further clarification to improve the accuracy of the diagnosis.

The experts in implant dentistry were the most accurate, showing substantial agreement with the gold standard diagnosis (mean quadratic weighted kappa value of 0.646) and higher percentages of complete agreement with reference case definitions than the other groups ($p < 0.001$). Conversely, undergraduate students and general dentists would appear to benefit most from additional education

19433670, 0, Downloaded from https://aap.onlinelibrary.wiley.com/doi/10.1002/JPER.23-0129 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and training. One of the undergraduate dental students (#8) achieved a low negative kappa value (−0.007), which was interpreted as "no agreement." Moreover, it should be noted that large negative kappa values could lead to erroneous interpretations, results, and conclusions.[23]

In this study, accuracy was also affected by the absence of longitudinal data. In fact, in cases without baseline readings examiners were statistically significantly less accurate ($p < 0.001$), confirming the importance of baseline readings in the diagnosis of peri-implant diseases.[24] However, although cases with and without baseline readings were included in this study in order to provide the examiner with a more realistic situation in which the clinician is confronted with both scenarios, these results should be considered as exploratory because of the limited number of dental implants for each type of case.

Concerning the consistency of diagnosis, the inter-examiner agreement was moderate for all groups of raters (experts, k = 0.52; general dentists, k = 0.53; and undergraduate students, k = 0.46), both in cases with and without baseline readings (in the presence of longitudinal data, k = 0.48; in the absence of longitudinal data, k = 0.47). Merli et al. previously assessed the inter-rater agreement in the diagnosis of peri-implant disease according to the definition of the 7th European Workshop on Periodontology.[25,26] In their study, the agreement between three experienced examiners was substantial (Fleiss k-statistic with square weight was 0.66). The higher reliability with respect to the inter-examiner agreement obtained by the dental implant experts participating in the present survey could be justified not only by the differences in the case definitions (7th European Workshop on Periodontology vs. 2017 World Workshop), but also by the different number of observers and the method of examining the dental implants (clinical examination vs. evaluation of clinical and radiographic data).

4.2 | Challenges in the determination of case definition

Using the diagnostic criteria of the 2018 classification of peri-implant health and disease case definitions, some inconsistencies have emerged in the presence of specific clinical conditions, even among experts. In particular, the analysis of the results of this panel of examiners was invaluable in highlighting the gray zones discussed in this section where all participants showed reduced agreement with the gold standard diagnosis.

A frequent clinical scenario has been the presence of single or very limited positive sites for bleeding on probing in otherwise healthy implant cases. Strictly following the guidelines, the presence of BoP can change the diagnosis from a healthy implant to mucositis. However, bleeding at implant sites could often be the result of trauma rather than inflammation, due to the mechanical fragility of the peri-implant tissues. In addition to variations in probing (amount of applied force, type of probe, and technique), it could be influenced by various factors such as peri-implant biofilm, host-related factors, and the design of the implant-supported prosthesis.[27] Furthermore, although the presence of BoP at the implant site is associated with a high negative predictive value and high sensitivity, 3 to 4 positive BoP sites were found to be the strongest predictor of advanced disease progression.[28–30] Nevertheless, non-dichotomous scales are recommended for classifying BoP to improve accuracy in diagnosing inflammatory conditions (e.g., mucositis or peri-implantitis) and better discriminate between profuse bleeding (as required by case definitions) and bleeding spots.[31] In the present study, otherwise healthy cases exhibiting only 1 to 2 sites positive for bleeding were incorrectly considered healthy by the examiners in 77.8% of cases, underestimating the presence of peri-implant mucositis (e.g., cases number 8, 12, and 23). In this regard, an international initiative aimed at developing a core outcome set and measurements for implant dentistry clinical trials (ID-COSM) recently recommended to discriminate between implants with a limited extent of BoP (≤1 spot/implant, not line or profuse bleeding)—which should be considered acceptable—and implants with extensive BoP (≥2 spots/implant or ≥1 site/implant with a line or profuse bleeding).[6]

Furthermore, in the present study, positive cases for PD of ≥6 mm and bleeding, but exhibiting bone level equal to 1 or 2 mm, in the absence of longitudinal data, led to inconsistencies in the diagnosis due to the difficulty in discriminating between mucositis and peri-implantitis (e.g., cases numbers 4, 10, 15 vs. 21, 22). In particular, implants without baseline readings and with the aforementioned clinical and radiographic signs were wrongly diagnosed as having peri-implantitis by 65.5% of the examiners. However, in the absence of baseline readings, 3 mm of bone loss are required—along with PD of ≥6 mm and bleeding on probing—to define cases of peri-implantitis according to the 2018 classification. This threshold has been explained because it is generally perceived that after implant placement and initial loading, part of the crestal bone height is lost (between 0.5 and 2 mm) during the healing process. However, this definition showed low sensitivity, especially for the early/incipient forms.[24] Incipient cases could then be left undiagnosed and untreated beyond the time they would have needed less invasive treatment and show better long-term outcomes.[32] Therefore, a possible reduction in the threshold for bone level (e.g., from ≥3 mm to ≥2 mm) in the absence of longitudinal data could be considered.

19433670, 0, Downloaded from https://aap.onlinelibrary.wiley.com/doi/10.1002/JPER.23-0129 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

While not present within the sample of implants examined for this investigation, three further scenarios were possible, which could have impaired reliability and accuracy among examiners.

First, incongruities and inaccuracies could arise in implant cases exhibiting the presence of isolated clinical and/or radiographic signs in different implant sites. For example, this could include cases where an implant is positive for bleeding and negative for probing depth and bone loss in the mesio-buccal aspect and at the same time negative for bleeding and positive for probing depth and bone loss in the disto-palatal aspect. Indeed, these cases should theoretically be diagnosed as peri-implantitis.

Second, the absence of bleeding combined with the presence of one or both of the other parameters required for secondary case definition (probing depth ≥6 mm and bone level ≥3 mm) could have led to difficulties in assigning the correct case definition using indirect evidence. These situations, in fact, do not allow the diagnosis of peri-implantitis or mucositis and should be considered as peri-implant health in cases previously affected by peri-implantitis or with deep mucous tunnel.

Finally, in cases displaying gingival recessions, negative for PD ≥ 6 mm, but positive for BoP and bone level ≥ 3 mm, diagnosis of peri-implantitis in the absence of longitudinal data could not be made.

4.3 | Supplementary factors affecting reproducibility of peri-implant case definitions

The 2018 classification of peri-implant health and diseases encouraged a comprehensive examination including probing of peri-implant tissues and radiographic bone level analysis, which are influenced by reliable assessment.

Further information regarding the reproducibility of the PD[33–41] and BoP[26] measurement as well as the agreement in bone level assessment on intraoral radiographs[42–45] is available as Supplementary Material S3 in the online *Journal of Periodontology*.

4.4 | Limitations

Among the limitations of this research, it should be considered that the clinical and radiographic data to be evaluated to define each implant case were not collected by observers. Moreover, it should also be mentioned that probing was performed by a single member of the study team not involved in the reliability assessment without removing the implant restorations. However, the aim of the present study was to assess the consistency and accuracy in the assignment of the case definitions according to the 2018 classification rather than in the overall diagnostic process.

Furthermore, similar to comparable studies in this field, intra-rater agreement was not assessed along with inter-rater agreement.[10,20] Therefore it is suggested to integrate it in future investigations.

It should also be mentioned that no cases with adjacent implants and other types of superstructures were included. Moreover, the brand and characteristics of implants were not standardized. However, the latter aspect could have provided a more realistic representation of clinical practice.

Finally, although in the agreement studies it is recommended to determine the number of examiners based on generalizability and feasibility,[46] further studies on this topic could also include a larger number of examiners. Hence, the present research could represent a pilot to test the accuracy and reproducibility among cross-cultural clinicians on the diagnosis of peri-implant tissue health status. Such participants to be included in future surveys could also be randomly selected to account for bias.

5 | CONCLUSIONS

In summary, both the reliability and accuracy in assigning the case definitions of peri-implant health, peri-implant mucositis, and peri-implantitis according to the 2018 classification were mostly moderate. Complete agreement with the gold standard diagnosis was achieved in just over half of the cases and was unfavorably affected by the absence of longitudinal data and the lack of advanced education and training in implant dentistry. Proper interpretation of the presence of isolated sites with BoP/SoP was a key element in discriminating, respectively, between peri-implant health and peri-implant mucositis. Likewise, the correct assessment of the radiographic bone level was of paramount importance in distinguishing between peri-implant mucositis and peri-implantitis.

The practical implication of this study may be to consider refining the peri-implant tissue health case definition criteria based on the aforementioned findings, while strengthening their understanding by clinicians and researchers.

AUTHOR CONTRIBUTIONS

Lorenzo Marini: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Visualization, Writing—Original Draft Preparation, Writing—Review & Editing. Cristiano Tomasi: Investigation, Formal Analysis, Writing—Original Draft Preparation, Writing—Review & Editing. Jan Derks: Investigation, Formal Analysis,

19433670, 0, Downloaded from https://aap.onlinelibrary.wiley.com/doi/10.1002/JPER.23-0129 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Writing—Original Draft Preparation, Writing—Review & Editing. Filippo Graziani: Investigation, Writing—Review & Editing. Luca Landi: Investigation, Writing—Review & Editing. Mauro Merli: Investigation, Writing—Review & Editing. Luigi Nibali: Investigation, Writing—Review & Editing. Mario Roccuzzo: Conceptualization, Investigation, Writing—Review & Editing. Nicola M. Sforza: Conceptualization, Investigation, Writing—Review & Editing. Maurizio S. Tonetti: Investigation, Writing—Review & Editing. Federico Deli: Investigation, Resources, Writing—Review & Editing. Piero Papi: Resources, Writing—Review & Editing. Bianca Di Murro: Resources, Writing—Review & Editing. Mariana A. Rojas: Resources, Writing—Review & Editing. Andrea Pilloni: Conceptualization, Investigation, Methodology, Supervision, Writing—Original Draft Preparation, Writing—Review & Editing.

ACKNOWLEDGMENTS

The authors are grateful to Prof. Jan Derks (University of Gothenburg, Gothenburg, Sweden) for his contribution in the role of gold standard examiner. The study was self-funded by the authors and their institutions.

CONFLICT OF INTEREST STATEMENT

The authors state that they have no conflicts of interest, financial, or commercial relationships to declare.

DATA AVAILABILITY STATEMENT

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

ORCID

Lorenzo Marini https://orcid.org/0000-0001-6662-2644
Cristiano Tomasi https://orcid.org/0000-0002-3610-6574
Filippo Graziani https://orcid.org/0000-0001-8780-7306
Luca Landi https://orcid.org/0000-0002-8548-4395
Luigi Nibali https://orcid.org/0000-0002-7750-5010
Mario Roccuzzo https://orcid.org/0000-0002-2135-6503
Maurizio S. Tonetti https://orcid.org/0000-0002-2743-0137
Piero Papi https://orcid.org/0000-0003-2564-530X
Bianca Di Murro https://orcid.org/0000-0002-5342-5446
Mariana A. Rojas https://orcid.org/0000-0002-8712-0958
Andrea Pilloni https://orcid.org/0000-0002-6268-1863

REFERENCES

1. Schwarz F, Derks J, Monje A, Wang HL. Peri-implantitis. *J Periodontol*. 2018;89(Suppl 1):S267-S290.
2. Heitz-Mayfield LJA, Salvi GE. Peri-implant mucositis. *J Clin Periodontol*. 2018;45(Suppl 20):S237-S245.
3. Berglundh T, Armitage G, Araujo MG, et al. Peri-implant diseases and conditions: consensus report of workgroup 4 of the 2017 world workshop on the classification of periodontal and peri-implant diseases and conditions. *J Clin Periodontol*. 2018;45(Suppl 20):S286-S291.
4. Renvert S, Persson GR, Pirih FQ, Camargo PM. Peri-implant health, peri-implant mucositis, and peri-implantitis: case definitions and diagnostic considerations. *J Periodontol*. 2018;89(Suppl 1):S304-S312.
5. Ramanauskaite A, Daugela P, Juodzbalys G. Treatment of peri-implantitis: meta-analysis of findings in a systematic literature review and novel protocol proposal. *Quintessence Int*. 2016;47:379-393.
6. Tonetti MS, Sanz M, Avila-Ortiz G, et al. Relevant domains, core outcome sets and measurements for implant dentistry clinical trials: The Implant Dentistry Core Outcome Set and Measurement (ID-COSM) international consensus report. *J Clin Periodontol*. Published online May 4, 2023. doi:10.1111/jcpe.13808
7. Herrera D, Berglundh T, Schwarz F, et al. Prevention and treatment of peri-implant diseases – the EFP S3 level clinical practice guideline. *J Clin Periodontol*. 2023. accepted for publication.
8. Murphy EA. *The Logic of Medicine*. 2nd ed. The Johns. Hopkins University Press; 1997:119-136.
9. Marini L, Tonetti MS, Nibali L, et al. The staging and grading system in defining periodontitis cases: consistency and accuracy amongst periodontal experts, general dentists and undergraduate students. *J Clin Periodontol*. 2021;48:205-215.
10. Ravidà A, Travan S, Saleh MHA, et al. Agreement among international periodontal experts using the 2017 world workshop classification of periodontitis. *J Periodontol*. 2021;92:1675-1686.
11. Abrahamian L, Pascual-LaRocca A, Barallat L, et al. Intra- and inter-examiner reliability in classifying periodontitis according to the 2018 classification of periodontal diseases. *J Clin Periodontol*. 2022;49:732-739.
12. Pini Prato G, Di Gianfilippo R, Pannuti CM, et al. Diagnostic reproducibility of the 2018 classification of gingival recession defects and gingival phenotype: a multicenter inter- and intra-examiner agreement study. *J Periodontol*. 2023;94(5):661-672.
13. Kottner J, Audigé L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol*. 2011;64:96-106.
14. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;70:213-220.
15. Fleiss JL. *Statistical methods for rates and proportions*. 2nd ed. John Wiley & Sons; 1981:211-236.
16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
17. Rotondi MA, Donner A. A confidence interval approach to sample size estimation for interobserver agreement studies with multiple raters and outcomes. *J Clin Epidemiol*. 2012;65:778-784.
18. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 4th ed. Oxford University Press; 2003:167-210.
19. Isaia F, Gyurko R, Roomian TC, Hawley CE. The root coverage esthetic score: intra-examiner reliability among dental students and dental faculty. *J Periodontol*. 2018;89:833-839.

19433670, 0, Downloaded from https://aap.onlinelibrary.wiley.com/doi/10.1002/JPER.23-0129 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

20. Barootchi S, Mancini L, Heck T, et al. Reliability assessment of the classification for facial peri-implant soft tissue dehiscence/deficiencies (PSTDs): a multi-center inter-rater agreement study of different skill-level practitioners. *J Periodontol*. 2022;93:1173-1182.
21. Pini-Prato G. The Miller classification of gingival recession: limits and drawbacks. *J Clin Periodontol*. 2011;38:243-245.
22. Hefti AF, Preshaw PM. Examiner alignment and assessment in clinical periodontal research. *Periodontol 2000*. 2012;59:41-60.
23. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22:276-282.
24. Romandini M, Berglundh J, Derks J, Sanz M, Berglundh T. Diagnosis of peri-implantitis in the absence of baseline data: a diagnostic accuracy study. *Clin Oral Implants Res*. 2021;32:297-313.
25. Lang NP, Berglundh T, Working Group 4 of Seventh European Workshop on Periodontology. Periimplant diseases: where are we now?–consensus of the seventh european workshop on periodontology. *J Clin Periodontol*. 2011;38(Suppl 11):178-181.
26. Merli M, Bernardelli F, Giulianelli E, et al. Inter-rater agreement in the diagnosis of mucositis and peri-implantitis. *J Clin Periodontol*. 2014;41:927-933.
27. Dukka H, Saleh MHA, Ravidà A, Greenwell H, Wang HL. Is bleeding on probing a reliable clinical indicator of peri-implant diseases? *J Periodontol*. 2021;92:1669-1674.
28. Jepsen S, Rühling A, Jepsen K, Ohlenbusch B, Albers HK. Progressive peri-implantitis. Incidence and prediction of peri-implant attachment loss. *Clin Oral Implants Res*. 1996;7:133-142.
29. Luterbacher S, Mayfield L, Brägger U, Lang NP. Diagnostic characteristics of clinical and microbiological tests for monitoring periodontal and peri-implant mucosal tissue conditions during supportive periodontal therapy (SPT). *Clin Oral Implants Res*. 2000;11:521-529.
30. Karlsson K, Derks J, Håkansson J, Wennström JL, Petzold M, Berglundh T. Interventions for peri-implantitis and their effects on further bone loss: a retrospective analysis of a registry-based cohort. *J Clin Periodontol*. 2019;46:872-879.
31. Monje A, Amerio E, Farina R, et al. Significance of probing for monitoring peri-implant diseases. *Int J Oral Implantol (Berl)*. 2021;14:385-399.
32. Figuero E, Graziani F, Sanz I, Herrera D, Sanz M. Management of peri-implant mucositis and peri-implantitis. *Periodontol 2000*. 2014;66:255-273.
33. Caram SJ, Huynh-Ba G, Schoolfield JD, Jones AA, Cochran DL, Belser UC. Biologic width around different implant-abutment interface configurations. A radiographic evaluation of the effect of horizontal offset and concave abutment profile in the canine mandible. *Int J Oral Maxillofac Implants*. 2014;29:1114-1122.
34. Hermann JS, Buser D, Schenk RK, Schoolfield JD, Cochran DL. Biologic width around one- and two-piece titanium implants. *Clin Oral Implants Res*. 2001;12:559-571.
35. Schou S, Holmstrup P, Stoltze K, Hjørting-Hansen E, Fiehn NE, Skovgaard LT. Probing around implants and teeth with healthy or inflamed peri-implant mucosa/gingiva. A histologic comparison in cynomolgus monkeys (Macaca fascicularis). *Clin Oral Implants Res*. 2002;13:113-126.
36. Lang NP, Wetzel AC, Stich H, Caffesse RG. Histologic probe penetration in healthy and inflamed peri-implant tissues. *Clin Oral Implants Res*. 1994;5:191-201.
37. Cha J, Wadhwani C, Wang M, Hokett SD, Katancik J. Instrument selection and application used to probe dental implants. *Int J Oral Maxillofac Implants*. 2019;34:115-123.
38. Serino G, Turri A, Lang NP. Probing at implants with peri-implantitis and its relation to clinical peri-implant bone loss. *Clin Oral Implants Res*. 2013;24:91-95.
39. Koldsland OC, Scheie AA, Aass AM. Prevalence of peri-implantitis related to severity of the disease with different degrees of bone loss. *J Periodontol*. 2010;81:231-238.
40. Di Gianfilippo R, Askar H, Henderson J, et al. Intra- and interexaminer repeatability of diagnostic peri-Implant clinical measurement: a pilot study. *J Oral Implantol*. 2022;48:295-300.
41. Ramanauskaite A, Obreja K, Schwarz F, et al. Reliability of probing depth assessments at healthy implant sites and natural teeth. *Clin Oral Investig*. Published December 3, 2022. doi:10.1007/s00784-022-04810-5
42. Cortellini P, Cortellini S, Bonaccini D, Tonetti MS. Papilla preservation and minimally invasive surgery for the treatment of peri-implant osseous defects. Clinical and radiographic outcomes of a 5-year retrospective study. *Clin Oral Implants Res*. 2021;32:1384-1396.
43. Kullman L, Al-Asfour A, Zetterqvist L, Andersson L. Comparison of radiographic bone height assessments in panoramic and intraoral radiographs of implant patients. *Int J Oral Maxillofac Implants*. 2007;22:96-100.
44. Cassetta M, Di Giorgio R, Barbato E. Are intraoral radiographs reliable in determining peri-implant marginal bone level changes? The correlation between open surgical measurements and peri-apical radiographs. *Int J Oral Maxillofac Surg*. 2018;47:1358-1364.
45. Zhang CN, Zhu Y, Fan LF, Zhang X, Jiang YH, Gu YX. Intra- and inter-observer agreements in detecting peri-implant bone defects between periapical radiography and cone beam computed tomography: a clinical study. *J Dent Sci*. 2021;16:948-956.
46. Karanicolas PJ, Bhandari M, Kreder H, et al. Evaluating agreement: conducting a reliability study. *J Bone Joint Surg Am*. 2009;91(Suppl 3):99-106.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Marini L, Tomasi C, Gianserra R, et al. Reliability assessment of the 2018 classification case definitions of peri-implant health, peri-implant mucositis, and peri-implantitis. *J Periodontol*. 2023;1-14. https://doi.org/10.1002/JPER.23-0129

19433670, 0, Downloaded from https://aap.onlinelibrary.wiley.com/doi/10.1002/JPER.23-0129 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License