Received: 19 August 2022 | Revised: 15 October 2022 | Accepted: 8 November 2022

DOI: 10.1002/JPER.22-0501

JOURNAL OF Periodontology

ORIGINAL ARTICLE

Diagnostic reproducibility of the 2018 Classification of Gingival Recession Defects and Gingival Phenotype: A multicenter inter- and intra-examiner agreement study

Giovanpaolo Pini Prato[1] | Riccardo Di Gianfilippo[2] | Claudio M. Pannuti[3] | Edward P. Allen[4] | Sofia Aroca[5] | Gustavo Avila-Ortiz[6,7,8] | Philippe Bouchard[9] | Francesco Cairo[10] | Tali Chackartchi[11] | Pierpaolo Cortellini[12,13,14] | Debora Franceschi[15] | Ajay Mahajan[16,17] | Evelyn A. Mancini[18] | Michael K. McGuire[19] | Tea Natsvlishvili[20,21] | Mauro P. Santamaria[22,23] | E. Todd Scheyer[19] | Hom-Lay Wang[2] | Kenneth Kornman[2] | Leandro Chambrone[24,25,26]

[1]Tuscany Academy of Dental Research (ATRO), Florence, Italy

[2]Department of Periodontics and Oral Medicine, The University of Michigan—School of Dentistry, Ann Arbor, Michigan, USA

[3]Discipline of Periodontics, School of Dentistry, University of São Paulo, São Paulo, Brazil

[4]Private practice, Dallas, Texas, USA

[5]26K Center for Clinical Research, Paris, France

[6]Private Practice, Gonzalez Solano Atelier Dental, Madrid, Spain

[7]Department of Oral Medicine, Infection, and Immunity, Harvard School of Dental Medicine, Boston, Massachusetts, USA

[8]Department of Periodontics, University of Iowa College of Dentistry, Iowa City, Iowa, USA

[9]Department of Periodontology and URP2496, Université Paris Cité, Paris, France

[10]Head Research Unit in Periodontology and Periodontal Medicine, The University of Florence, Florence, Italy

[11]Department of Periodontology, Hadassah Medical Center, Faculty of Dental Medicine, Hebrew University of Jerusalem, Jerusalem, Israel

[12]Private practice, Florence, Italy

[13]European Research Group on Periodontology (ERGOPERIO), Genova, Italy

[14]Deptartment of Oral Health Sciences, KU Leuven & Dentistry (Periodontology), Un Hospitals Leuven, Leuven, Belgium

[15]Department of Experimental and Clinical Medicine, The University of Florence, Florence, Italy

[16]H.P.Governmental Dental College and Hospital, Shimla, Himachal Pradesh, India

[17]King's College, London, UK

[18]Private practice, Rosario, Argentina

[19]Perio Health Professionals, The McGuire Institute, Houston, Texas, USA

[20]Albius International Perio School, Tbilisi, Georgia

[21]Section of Periodontology, Department of Oral and Maxillofacial Sciences, The University of Rome, la Sapienza, Rome, Italy

[22]Division of Periodontics, Institute of Science and Technology, São José dos Campos, São Paulo State University (UNESP), São Paulo, Brazil

[23]College of Dentistry, University of Kentucky, Lexington, Kentucky, USA

[24]Evidence-Based Hub, Centro de Investigação Interdisciplinar Egas Moniz (CiiEM), Egas Moniz-Cooperativa de Ensino Superior, Almada, Portugal

[25]Department of Periodontics, The University of Pennsylvania, Philadelphia, Pennsylvania, USA

[26]Unit of Basic Oral Investigation (UIBO), Universidad El Bosque, Bogota, Colombia

Correspondence
Giovanpaolo Pini Prato, Tuscany Academy of Dental Research (ATRO), Viale Matteotti 11, Florence 50121, Italy.
Email: studio.pinipratoagmail.com

Abstract

Background: The aim of this study was to investigate the inter- and intra-examiner agreement among international experts on the diagnosis of gingival recession defects using the 2018 Classification of Gingival Recession Defects and Gingival Phenotype as proposed in the 2017 World Workshop.

Methods: Standardized intraoral photographs from 28 gingival recession defects were evaluated twice by 16 expert periodontists. Recession type (RT), recession depth (RD), keratinized tissue width (KTW), gingival thickness (GT), detectability of the cemento-enamel junction (CEJ), and presence of root steps (RS) were recorded and used for the analysis. Intra- and inter-examiner agreements were calculated for individual variables and for the overall classification. Intraclass correlation coefficient with 95% CI was used for RD and KTW; Kappa with 95% CI was used for GT, CEJ, and RS; quadratic weighted Kappa with 95% CI was used for RT.

Results: Overall intra- and inter-examiner agreements were highest for KTW (0.95 and 0.90), lowest for GT (0.75 and 0.41), with the other variables in between (RD: 0.93 and 0.68, RS: 0.87 and 0.65, RT: 0.79 and 0.64, CEJ: 0.75 and 0.57). Overall intra- and inter-examiner agreements for the matrix were 62% and 28%, respectively. Significant effects existed between one variable's measurement and other variables' agreements.

Conclusions: The 2018 Classification of Gingival Recession Defects and Gingival Phenotype is clinically reproduceable within the examiners, and when the variables forming the matrix are analyzed individually. The between-examiner agreement for the complete matrix showed lower reproducibility. The agreement was highest for KTW and RD, and least for GT.

KEYWORDS
classification, connective tissue, gingiva, gingival recession, periodontics, phenotype, tooth root

1 | INTRODUCTION

Diagnostic classifications are systematic arrangements into categories according to established criteria; they are indispensable in the medical fields for reproducible diagnosis and subsequent treatment planning.[1] In periodontal plastic surgery, correct diagnosis of gingival recession defects (GRDs) is of crucial importance to define the need for therapy, to choose among suitable treatment approaches, and to monitor or predict longitudinal changes of treated versus untreated sites. Standardized and precise diagnosis is also a prerequisite for effective communication among multiple providers and for homogeneity of data for research purposes.

Many diagnostic classifications of GRDs have been proposed over the years.[2–8] Among them, Miller's classification system has been, since its publication in 1985, the most emblematic in research and clinical practice for decades.[4] However, this classification system considers a limited set of clinical variables (i.e., characteristics of the gingiva at the site affected by a GRD) and does not account for other relevant features, such as the root-surface characteristics.[4]

The creation of a new classification may be a challenging task.[1] Ideally, the final product should satisfy the characteristics of suitable taxonomy described by Murphy.[9] A classification should prove usefulness by being applicable for clinical use. It should be simultaneously exhaustive and non-ambiguous, with each possible entity accommodated into one and only one specific class or category. Finally, it should be of simple use and with high reproducibility.[9,10]

To overcome the limitations of previous systems, a new classification for GRDs was proposed and approved during the 2017 World Workshop on Periodontal and Peri-implant Diseases and Conditions, and was subsequently published in 2018.[7] This classification, generally known as the 2018 Classification of Gingival Recession Defects and Gingival Phenotype, consists of a 4 × 5 matrix that integrates the most relevant clinical variables from previous systems including: (i) recession defect-related features, such as interproximal recession type (RT)[6] and midfacial recession depth (RD); (ii) local phenotypic characteristics, such as gingival thickness (GT) and keratinized tissue width (KTW); and (iii) and root surface attributes, such as detectability of the cemento-enamel junction (CEJ) and presence of non-carious cervical lesions.

Testing its external validity was encouraged in the original publication of this new classification.[7] Recent research has already been conducted to explore the epidemiology of the new classification classes.[11] Additionally, commentaries have been published to highlight its features and to instruct on its use in clinical and research settings.[10] However, to the best of our knowledge, no study has investigated the diagnostic reproducibility of the new classification. Furthermore, no information on the agreement among periodontists using the new matrix is available.

The use of photographs as a form of diagnostic imaging has gained popularity for studies on measurement repeatability and operator agreement. Studies have tested the applicability of photographs to investigate gingival esthetics[11-13] but, within our knowledge, studies on the diagnostic use of photographs never targeted the 2018 Classification of Gingival Recession Defects and Gingival Phenotype. Therefore, the aim of this multicenter study was to use a photographic model to investigate the inter- and intra-examiner agreement on the assessment and categorization of GRDs using the 2018 Classification of Gingival Recession Defects and Gingival Phenotype among international experts.

2 | MATERIALS AND METHODS

Based on the features of the 2018 Classification of Gingival Recession Defects and Gingival Phenotype proposed by Cortellini and Bissada,[7] a sample of 28 single GRDs from 17 patients treated in a private office setting (Florence 50121, Italy) was selected for this inter- and intra-examiner reproducibility study. The sample consisted of both maxillary and mandibular teeth (nine incisors, nine cuspids, and 10 bicuspids). Due to the nature of this study, no ethical committee approval was required. All patients gave consent for the use of intraoral photos for research purposes. This study was conducted in accordance with the 1975 Helsinki Declaration as revised in 2000, 2008, and 2013.

2.1 | Inter- and intra-examiner assessment

A group of 16 periodontists, from Europe, North America, South America, and Asia, were invited to participate in the study, namely Drs. E.P. Allen, S. Aroca, G. Avila-Ortiz, P. Bouchard, F. Cairo, T. Chackartchi, P. Cortellini, R. Di Gianfilippo, D. Franceschi, A. Mahajan, E.A. Mancini, M.K. McGuire, T. Natsvlishvili, M.P. Santamaria, E.T. Scheyer, and H-L. Wang. These individuals were selected based on their clinical and/or research expertise on the treatment of GRDs. Upon agreeing to participate in this study, pre-established standard guidelines for the assessment of GRDs were sent to all examiners. The process of assessment and classification was performed in two phases.

Phase 1

A PDF document containing the 28 cases of GRDs selected for analysis was forwarded to the examiners. Each GRD was presented using seven standardized intraoral photographs (Figures 1 and 2). A combination of frontal view and lateral view with a periodontal probe[*] was used to aid on the assessment of the following parameters:

1. RT 1, 2, or 3 evaluated based on interproximal probing.[6]
2. RD in millimeters evaluated using the midfacial probe as a reference.
3. GT (thin vs. thick) evaluated by tissue transparency upon midfacial sulcus probing.
4. Keratinized tissue width (KTW) in millimeters measured using a midfacial probe running from the free gingival margin to the mucogingival junction.
5. Integrity of the CEJ (A vs. B) evaluated with a frontal and lateral view.[5]
6. Presence or absence of root steps (RS) deeper than 1 mm in the horizontal dimension (+ vs. -) evaluated with a lateral view.[5]

The examiners were instructed to round the measurements to the nearest 1.0 mm for RD and KTW. On the PDF file, the examiners populated a data collection sheet designed for this study to record information relative to all the previously outlined parameters. After classifying all sites, the answer sheet was submitted to the study's central registrar.

Phase 2

Two weeks after completion of phase I, all examiners received a second PDF document and a new blank data collection sheet. The new PDF document contained the same group of 28 cases already screened by the raters but sorted in another random order. Randomization of cases

[*] CP-15 UNC periodontal probe, Hu-Friedy, Chicago, Illinois.

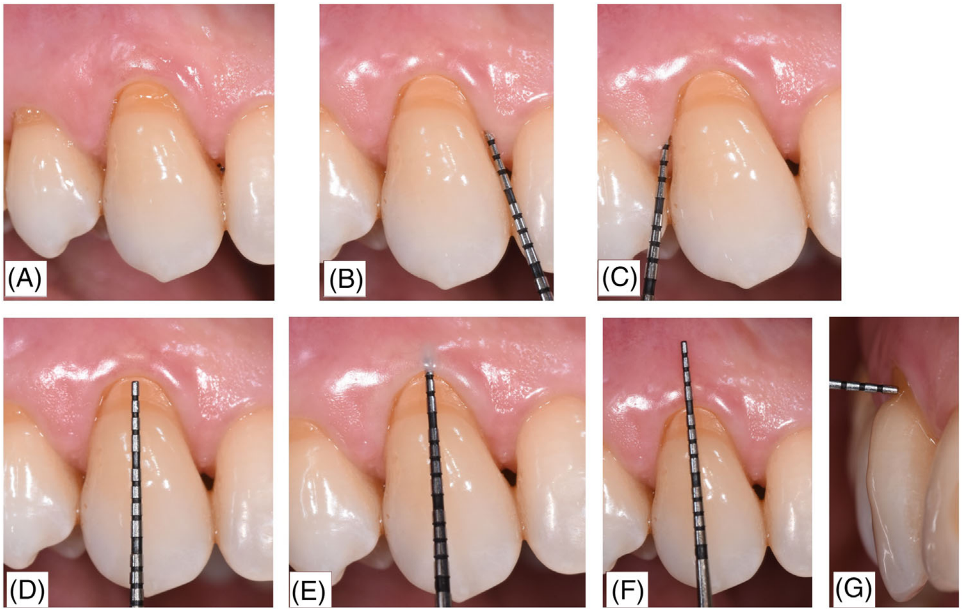

FIGURE 1 Photographic documentation of a maxillary right canine illustrated with a series of seven intraoral images from a frontal and lateral view, with and without a UNC-15 probe. Frontal image (**A**) was captured to provide an overview of the gingival recession defect. Mesial (**B**) and distal (**C**) interproximal probing were used for calculation of the recession type. Midfacial position of the UNC-15 probe was used to record the recession depth (**D**) measured from the cemento-enamel junction to the free gingival margin, gingival thickness (**E**) was evaluated by tissue translucency at midfacial probing, and keratinized tissue width (**F**) was measured from the free gingival margin to the mucogingival junction. Lateral view (**G**) of the gingival recession defect allowed measurement of the depth of the non-carious cervical lesion. Integrity of the cemento-enamel junction was assessed at frontal and lateral view.

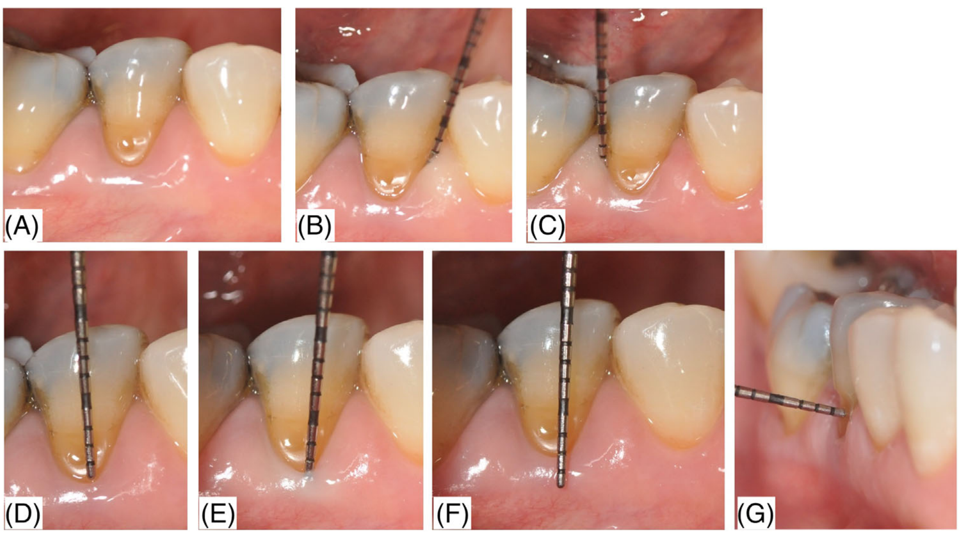

FIGURE 2 Photographic documentation of a mandibular right premolar illustrated with a series of seven intraoral images from a frontal and lateral view, with and without a UNC-15 probe. Frontal image (**A**) was captured to provide an overview of the gingival recession defect. Mesial (**B**) and distal (**C**) interproximal probing were used for calculation of the Recession Type. Midfacial position of the UNC-15 probe was used to record the recession depth (**D**) measured from the cemento-enamel junction to the free gingival margin, gingival thickness (**E**) was evaluated by tissue translucency at midfacial probing, and keratinized tissue width (**F**) was measured from the free gingival margin to the mucogingival junction. Lateral view (**G**) of the gingival recession defect allowed measurement of the depth of the non-carious cervical lesion. Integrity of the cemento-enamel junction was assessed at frontal and lateral view.

was performed with random/sort function using a specific software package.[†] As in phase I, the examiners were invited to classify each one of the GRDs again. In addition, the examiners were instructed not to look back to their previous assessments.

2.2 | Sample size calculation

The sample size for this study was calculated using the method described by Walter et al.[14] (i.e., agreement assessment by the calculation of intraclass correlation [p]). Taking into consideration 16 examiners, with a minimum level of agreement (p_0) between them of 0.80 and a p_1 (i.e., alternative hypothesis) of 0.90 (with $\alpha = 0.05$ and $\beta = 0.20$), at least 23 GRD were required.

2.3 | Statistical analysis

Intra- and inter-examiner agreement was calculated via intraclass correlation coefficients (ICC) with 95% confidence intervals (95% CI) for quantitative variables (RD and KTW). Kappa with 95% CI was used to calculate agreement for qualitative nominal variables (GT, CEJ, and RS), and quadratic weighted Kappa with 95% CI was used to calculate agreement for RT, an ordinal variable. For inter-examiner agreement, examiner #1 (F. Cairo) was considered the gold-standard. Kappa was also used to assess intra- and inter-examiner agreement of RD and KTW by allowing a tolerance of error of 1 mm for agreement between the two measurements.

Mann-Whitney test was used to verify whether there was a significant difference between intra-and inter-examiner agreement regarding PD and REC. Kruskal–Wallis test was used to assess the intra-and inter-examiner agreement difference regarding GT, CEJ, and RS. We also verified the overall inter- and intra-examiner agreement on the complete matrix of the new classification (a composite of the six clinical variables) using the Kappa coefficient. For this analysis, we allowed a tolerance error of 1 mm for RD and KTW, and we considered that there was agreement only if the examiner agreed on all six variables at the same time. We used the following categorization to interpret Kappa values: poor agreement (<0.00); slight agreement (0.00 to 0.20); fair agreement (0.21 to 0.40); moderate agreement (0.41 to 0.60); substantial agreement (0.61 to 0.80); and almost perfect agreement (0.81 to 1.00).[15] ICC values were categorized as: poor agreement (<0.5); moderate agreement (0.5 to 0.75); good agreement (0.75 to 0.9); and excellent agreement (>0.90).[16] Categorization of Kappa and ICC were displayed in heat maps.

The effect of variables on the intra- and inter-operator agreements were calculated with binary logistic regression for nominal dichotomous variables (RS, CEJ, and GT) and with ordinal logistic regression for ordinal (RT) or continuous (RD and KTW) variables. Multicollinearity was verified with the variance inflation factor. Outliers were verified by Cooks' distance and by the residuals (in the case of binary logistic regression). ANOVA with Wald test was conducted to verify the significance of the predictors (independent variables) and for the main independent variable, the coefficients of the standard deviation were reported, as well as the odds ratio (OR), that indicates increase or decrease in the chance of agreement, with the correspondent p value. Analyses were carried out in a dedicated software for statistical computing.[‡] The significance level of statistical tests was set at 5%.

3 | RESULTS

The results of the two sets of measurements performed by the 16 examiners are depicted below (Tables 1–4). Phase 1 was conducted between October and November 2021, whereas phase 2 was performed between November and December 2021.

Table 1 shows the intra-examiner agreement for all variables. Including the 16 different examiners, intra-examiner reproducibility of RD and KTW was calculated with ICC and ranged from 0.67 to 1.00 for RD, and from 0.63 to 1.00 for KTW. Intra-examiner agreement for GT, CEJ, and RS was calculated with Kappa coefficient and ranged from 0.27 to 1.00 (GT), 0.27 to 1.00 (CEJ), and 0.55 to 1.00 (RS) among the different examiners. Weighted Kappa was used to assess intra-examiner agreement for RT and ranged from 0.42 to 1.00 among the different examiners. Overall intra-examiner agreement was 0.93 (95% CI, 0.92–0.94) for RD; 0.95 (95% CI, 0.94–0.96) for KTW; 0.75 (95% CI, 0.68–0.81) for GT; 0.75 (95% CI, 0.69–0.81) for CEJ; 0.87 (95% CI, 0.82–0.92) for RS; and 0.79 (95% CI, 0.67–0.92) for RT. Intra-examiner agreement was not significantly different between RD and KTW ($p = 0.06$), nor between GT, CEJ, and RS ($p = 0.12$).

When allowing a tolerance of error of 1 mm for RD and KTW, intra-examiner agreement ranged from 0.86 to 1.00 for RD, and from 0.78 to 1.00 for KTW. Overall intra-examiner agreement with a tolerance of 1 mm was 0.97 and 0.98 for RD and KTW, respectively (Table 2).

The overall intra-examiner agreement for the complete matrix of the new classification (simultaneous agreement

[†] Microsoft Excel. Microsoft Office, Redmond, Washington.

[‡] The open software R, version 4.1.1. RStudio (2021-08-10)

TABLE 1 Intra-examiner agreement for recession depth (RD), keratinized tissue width (KTW), gingival thickness (GT), cemento-enamel junction detectability (CEJ), root step (RS) and recession type (RT) by operator

Examiner	RD	KTW	GT	CEJ	RS	RT
1	0.67 (0.38–0.83)	0.94 (0.88–0.97)	0.43 (0.06–0.79)	0.70 (0.47–0.92)	0.85 (0.66–1.00)	0.87 (0.72–1.00)
2	0.96 (0.93–0.98)	0.94 (0.88–0.97)	0.88 (0.67–1.00)	1.00 (1.00–1.00)	0.91 (0.75–1.00)	0.76 (0.38–1.00)
3	0.96 (0.93–0.98)	0.95 (0.90–0.97)	0.75 (0.48–1.00)	0.83 (0.59–1.00)	0.92 (0.78–1.00)	0.90 (0.74–1.00)
4	1.00 (1.00–1.00)	0.99 (0.98–0.99)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.92 (0.78–1.00)	1.00 (1.00–1.00)
5	0.93 (0.86–0.96)	0.97 (0.95–0.99)	0.76 (0.43–1.00)	0.45 (0.13–0.77)	0.79 (0.56–1.00)	0.49 (0.09–0.88)
6	0.92 (0.83–0.96)	0.97 (0.96–0.98)	0.83 (0.59–1.00)	0.64 (0.35–0.93)	0.79 (0.51–1.00)	1.00 (1.00–1.00)
7	0.94 (0.89–0.97)	0.99 (0.98–0.99)	1.00 (1.00–1.00)	0.46 (0.12–0.80)	0.89 (0.67–1.00)	0.73 (0.43–1.00)
8	0.93 (0.87–0.96)	0.97 (0.94–0.98)	1.00 (1.00–1.00)	0.36 (0.00–1.00)	0.84 (0.62–1.00)	0.71 (0.42–0.99)
9	0.94 (0.89–0.97)	0.88 (0.77–0.94)	0.65 (0.19–1.00)	0.29 (0.00–0.79)	0.55 (0.23–0.87)	0.71 (0.42–0.99)
10	0.92 (0.84–0.96)	0.96 (0.92–0.98)	0.84 (0.52–1.00)	0.93 (0.79–1.00)	0.92 (0.77–1.00)	0.87 (0.68–1.00)
11	0.93 (0.86–0.96)	0.99 (0.98–0.99)	0.89 (0.67–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.94 (0.69–1.00)
12	0.94 (0.89–0.97)	0.93 (0.87–0.96)	0.55 (0.24–0.87)	0.79 (0.51–1.00)	0.90 (0.71–1.00)	0.89 (0.71–1.00)
13	0.99 (0.98–0.99)	1.00 (1.00–1.00)	0.76 (0.43–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.96 (0.85–1.00)
14	0.89 (0.78–0.94)	0.96 (0.93–0.98)	0.43 (0.06–0.79)	0.85 (0.66–1.00)	0.70 (0.47–0.92)	0.42 (0.03–0.80)
15	0.94 (0.88–0.97)	0.63 (0.30–0.81)	0.27 (0.00–0.60)	0.70 (0.47–0.92)	1.00 (1.00–1.00)	0.71 (0.43–1.00)
16	0.87 (0.75–0.93)	0.94 (0.89–0.97)	0.48 (0.15–0.81)	0.27 (0.00–0.60)	0.84 (0.62–1.00)	0.42 (0.00–0.97)
Overall	0.93 (0.92–0.94)	0.95 (0.94–0.96)	0.75 (0.68–0.81)	0.75 (0.69–0.81)	0.87 (0.82–0.92)	0.79 (0.67–0.92)

Values are presented as ICC (95% CI) for RD and KTW, Kappa coefficient (95% CI) for GT, CEJ, and RS, and quadratic weighted Kappa (95% CI) for RT.
Abbreviations: CEJ, cemento-enamel junction; KTW, keratinized tissue width; GT, gingival thickness, RS, root step; RT, recession type.

of all six variables) when allowing a tolerance of error of 1 mm for RD and KTW was 61.6%, with a Kappa coefficient of 0.23 (95% CI, 0.14–0.32).

Table 3 shows the inter-examiner agreement. Considering the agreement of the gold standard with the other 15 examiners, reproducibility of RD and KTW was calculated via ICC and ranged from 0.58 to 0.85 for RD, and from 0.60 to 0.97 for KTW. Inter-examiner agreement for GT, CEJ, and RS was calculated with Kappa coefficient and ranged from 0.05 to 0.89 (GT), 0.10 to 0.92 (CEJ), and 0.40 to 0.85 (RS) between the gold standard and the other examiners. Weighted Kappa was used to assess inter-examiner agreement for RT, ranging from 0.42 to 0.79. Overall inter-examiner agreement was 0.68 (95% CI, 0.62–0.73) for RD and 0.90 (95% CI, 0.87–0.92) for KTW. Inter-examiner agreement was significantly higher for KTW than for RD ($p < 0.001$). Further, overall inter-examiner agreement was 0.41 (95% CI, 0.31–0.51) for GT; 0.57 (95% CI, 0.49–0.65) for CEJ; 0.65 (95% CI, 0.59–0.71) for RS; and 0.64 (95% CI, 0.52–0.76) for RT. Intra-examiner agreement was not significantly different between GT, CEJ, and RS ($p = 0.33$).

When allowing a tolerance of error of 1 mm for RD and KTW, inter-examiner agreement ranged from 0.59 to 0.93 for RD, and from 0.72 to 1.00 for KTW. Overall inter-examiner agreement with a tolerance of 1 mm was 0.78 (95% CI, 0.72–0.83) for RD and 0.96 (95% CI, 0.93–0.98) for KTW (Table 4).

The overall inter-examiner agreement for the complete matrix of the new classification (agreement of all six variables) when allowing a tolerance of error of 1 mm for RD and KTW was 27.9%, with a Kappa coefficient of -0.36 (95% CI, -0.26 to 0.45).

Figures S1 and S2 depict the heat map of intra and inter-examiner agreement for the six clinical variables (see Figures S1 and S2 in online *Journal of Periodontology*).

The effect of a variable's measurement on the other variables' agreement was tested and marked as positive for increased chance of agreement, or negative for decreased chance of agreement. In the case of ordinal logistic regression, the inverse function was used as reported in the Tables S1 and S2 (see Tables S1 and S2 in online *Journal of Periodontology*).

A significant effect was noted for RD (mm) on the intra-operator agreement of RT (positive $p = 0.023$) and GT (positive, $p = 0.032$); for RT on the intra-operator agreement of GT (negative, $p < 0.001$); for GT on the intra-operator agreement of KTW (positive, $p = 0.009$); for detectability of CEJ on the intra-operator agreement of RD (positive, $p < 0.001$); and for detectability of CEJ on the intra-operator agreement of RT (positive, $p = 0.018$) (see Table S1 in online *Journal of Periodontology*). Further, there was an effect of the operator in all cases, except when testing the effect of CEJ in the agreement of RS, and in the effect of RS in the agreement of CEJ.

TABLE 2 Intra-examiner agreement for recession depth and keratinized tissue width when allowing a tolerance of error of 1 mm between the two measurements.

Examiner	RD	KTW
1	0.86 (0.66–1.00)	1.00 (1.00–1.00)
2	1.00 (1.00–1.00)	1.00 (1.00–1.00)
3	1.00 (1.00–1.00)	0.93 (0.79–1.00)
4	1.00 (1.00–1.00)	1.00 (1.00–1.00)
5	1.00 (1.00–1.00)	1.00 (1.00–1.00)
6	1.00 (1.00–1.00)	1.00 (1.00–1.00)
7	1.00 (1.00–1.00)	1.00 (1.00–1.00)
8	0.93 (0.79–1.00)	1.00 (1.00–1.00)
9	1.00 (1.00–1.00)	1.00 (1.00–1.00)
10	0.93 (0.79–1.00)	1.00 (1.00–1.00)
11	1.00 (1.00–1.00)	1.00 (1.00–1.00)
12	1.00 (1.00–1.00)	1.00 (1.00–1.00)
13	1.00 (1.00–1.00)	1.00 (1.00–1.00)
14	0.93 (0.79–1.00)	1.00 (1.00–1.00)
15	1.00 (1.00–1.00)	0.78 (0.56–1.00)
16	0.86 (0.66–1.00)	0.93 (0.79–1.00)
Overall	0.97 (0.95–0.99)	0.98 (0.96–0.99)

Values are presented as Kappa coefficient (95% CI).
Abbreviations: KTW, keratinized tissue width, RT, recession type.

A statistically significant effect was noted for RD (mm) on the inter-operator agreement of RT (positive, $p < 0.001$); for KTW on the inter-operator agreement of RD (positive, $p = 0.007$); for RT on the inter-operator agreement of RD (positive, $p < 0.001$) and GT (negative, $p = 0.033$); for GT on the inter-operator agreement of KTW (positive, $p = 0.02$); for detectability of CEJ on the inter-operator agreement of RD (positive, $p < 0.001$), RT (positive, $p = 0.001$), and RS (positive, $p = 0.037$); and for RS on the inter-operator agreement of RT (negative, $p = 0.004$), RD (negative, $p = 0.001$), and CEJ (negative, $p = 0.021$) (see Table S2 in online *Journal of Periodontology*). Further, there was an effect of the operator in all cases, except when testing the effect of CEJ in the agreement of RT and RS, and in the effect of RS in the agreement of CEJ.

4 | DISCUSSION

To our knowledge, this is the first study to assess the inter- and intra-examiner agreement (i.e., reproducibility) for the categorization of GRDs using the 2018 Classification of Gingival Recession Defects and Gingival Phenotype,[7] based on the perspective of 16 international expert periodontists.

4.1 | Intra- and inter-examiner agreement

With respect to the intra-examiner agreement assessments of all 16 examiners, these ranged from moderate to excellent/almost perfect for RD, KTW, RS, and RT. In contrast, reproducibility varied widely for GT and CEJ, with a range from fair to almost perfect. The overall intra-examiner agreement varied from substantial (GT, CEJ, and RT) to almost perfect (RD, KTW, and RS), but without significant differences between these variables ($p >$ 0.05). With respect to the inter-examiner agreement assessments, when the gold-standard examiner was compared with the other 15 raters, these ranged from slight to almost perfect for GT and CEJ, fair to almost perfect for RS, moderate to good/substantial for RD and RT and moderate to excellent for KTW. The overall inter-examiner agreement varied from moderate (GT, RD, CEJ) to good (RS, KTW, and RT). A superior level of inter-examiner agreement was detected for KTW compared with RD ($p < 0.001$). It has to be mentioned that RT was estimated based on interproximal probing depth and visual assessment of the papilla height in relation to the adjacent crowns. Probing depth could have been misleading for determining clinical attachment level when the gingival margin was coronal to the CEJ, and this could have negatively impacted the inter-examiner agreement of RT. Despite that, the repeatability of RT was substantial for both intra- and inter-examiner agreements (0.79 and 0.64, respectively).

Interestingly, statistically significant effects existed between variable measurements and agreements. Among the most notable, detectability of the CEJ improved the agreements for RD (both intra- and inter-). This stresses the importance of a detectable CEJ for accurate measurement of RD and its root-coverage surrogates. Also, detectability of CEJ positively related with agreements for RT (both intra- and inter-) that is an important variable to consider during the screening/enrollment phases of clinical studies as well as for communication of realistic expectations of root coverage in clinical settings. The presence of RS negatively affected the agreement of RD and CEJ. This was expected considering that a noncarious cervical lesion damages the surface of a midbuccal recession and often changes the anatomy of the CEJ. GT positively related with agreements of KTW (both intra- and inter-) so that thicker gingiva allowed clearer identification and more repeatable measurement of the anatomical landmarks used for measurement of KTW. Similarly, KTW affected the inter-examiner repeatability of RD as they both share the free gingival margin as a common anatomical landmark. Double-arrow positive interaction was noted for recession-related variables

TABLE 3 Inter-examiner agreement for recession depth (RD), keratinized tissue width (KTW), gingival thickness (GT), cemento-enamel junction detectability (CEJ), root step (RS), and recession type (RT) by gold standard (examiner #1) versus other examiners.

Examiner	RD	KTW	GT	CEJ	RS	RT
1 vs. 2	0.85 (0.71–0.92)	0.60 (0.29–0.81)	0.58 (0.19–0.96)	0.68 (0.39–0.97)	0.70 (0.44–0.97)	0.54 (0.21–0.87)
1 vs. 3	0.72 (0.47–0.85)	0.96 (0.92–0.98)	0.49 (0.12–0.85)	0.68 (0.39–0.97)	0.85 (0.66–1.00)	0.75 (0.53–0.97)
1 vs. 4	0.77 (0.57–0.88)	0.97 (0.94–0.98)	0.89 (0.67–1.00)	0.84 (0.63–1.00)	0.78 (0.54–1.00)	0.79 (0.60–0.98)
1 vs. 5	0.62 (0.18–0.82)	0.95 (0.90–0.97)	0.66 (0.29–1.00)	0.51 (0.22–0.80)	0.50 (0.18–0.82)	0.54 (0.08–0.99)
1 vs. 6	0.71 (0.40–0.85)	0.95 (0.91–0.98)	0.70 (0.37–1.00)	0.56 (0.25–0.97)	0.79 (0.24–0.87)	0.54 (0.08–0.99)
1 vs. 7	0.78 (0.58–0.88)	0.96 (0.92–0.98)	0.79 (0.50–1.00)	0.36 (0.00–0.73)	0.40 (0.05–0.75)	0.71 (0.41–1.00)
1 vs. 8	0.66 (0.35–0.82)	0.94 (0.86–0.97)	0.40 (0.05–0.75)	0.10 (0.00–0.55)	0.78 (0.54–1.00)	0.75 (0.52–0.98)
1 vs. 9	0.77 (0.57–0.88)	0.63 (0.30–0.81)	0.49 (0.12–0.85)	0.51 (0.22–0.80)	0.68 (0.39–0.97)	0.54 (0.21–0.87)
1 vs 10	0.72 (0.47–0.85)	0.89 (0.80–0.94)	0.61 (0.19–1.00)	0.71 (0.44–0.97)	0.85 (0.66–1.00)	0.75 (0.52–0.98)
1 vs. 11	0.66 (0.37–0.82)	0.94 (0.89–0.97)	0.66 (0.29–1.00)	0.92 (0.78–1.00)	0.56 (0.24–0.87)	0.59 (0.19–1.00)
1 vs. 12	0.74 (0.51–0.86)	0.88 (0.77–0.94)	0.37 (0.01–0.74)	0.56 (0.26–0.97)	0.48 (0.14–0.81)	0.76 (0.56–0.96)
1 vs. 13	0.75 (0.52–0.87)	0.95 (0.91–0.97)	0.75 (0.43–1.00)	0.70 (0.42–0.97)	0.63 (0.33–0.92)	0.60 (0.31–0.89)
1 vs. 14	0.66 (0.36–0.82)	0.88 (0.63–0.95)	0.58 (0.19–0.96)	0.77 (0.53–1.00)	0.56 (0.24–0.87)	0.61 (0.21–1.00)
1 vs. 15	0.59 (0.09–0.79)	0.74 (0.39–0.87)	0.05 (0.00–0.33)	0.76 (0.38–0.93)	0.71 (0.44–0.97)	0.68 (0.39–0.98)
1 vs. 16	0.58 (0.20–0.78)	0.96 (0.92–0.98)	0.43 (0.06–0.79)	0.17 (0.00–0.58)	0.71 (0.44–0.97)	0.42 (0.12–0.72)
Overall	0.68 (0.62–0.73)	0.90 (0.87–0.92)	0.41 (0.31–0.51)	0.57 (0.49–0.65)	0.65 (0.59–0.71)	0.64 (0.52–0.76)

Values were presented as ICC (95% CI) for RD and KTW, Kappa coefficient for GT, CEJ, and RS, and quadratic weighted Kappa for RT.
Abbreviations: CEJ, cemento-enamel junction; KTW, keratinized tissue width; GT, gingival thickness, RS, root step; RT, recession type.

TABLE 4 Inter-examiner agreement for recession depth and keratinized tissue width when allowing a tolerance of error of 1 mm between the two measurements

Examiner	RD	KTW
1 vs. 2	0.93 (0.79–1.00)	0.72 (0.46–0.97)
1 vs. 3	0.79 (0.56–1.00)	1.00 (1.00–1.00)
1 vs. 4	0.86 (0.66–1.00)	1.00 (1.00–1.00)
1 vs. 5	0.72 (0.47–0.97)	1.00 (1.00–1.00)
1 vs. 6	0.79 (0.56–1.00)	1.00 (1.00–1.00)
1 vs. 7	0.93 (0.79–1.00)	1.00 (1.00–1.00)
1 vs. 8	0.71 (0.45–0.97)	1.00 (1.00–1.00)
1 vs. 9	0.72 (0.46–0.97)	1.00 (1.00–1.00)
1 vs. 10	0.84 (0.64–1.00)	0.93 (0.79–1.00)
1 vs. 11	0.72 (0.46–0.97)	1.00 (1.00–1.00)
1 vs. 12	0.85 (0.66–1.00)	0.93 (0.79–1.00)
1 vs. 13	0.85 (0.66–1.00)	1.00 (1.00–1.00)
1 vs. 14	0.71 (0.45–0.97)	0.79 (0.57–1.00)
1 vs. 15	0.59 (0.30–0.89)	1.00 (1.00–1.00)
1 vs. 16	0.71 (0.45–0.97)	1.00 (1.00–1.00)
Overall	0.78 (0.72–0.83)	0.96 (0.93–0.98)

Values are presented as Kappa coefficient (95% CI).
Abbreviations: KTW, keratinized tissue width, RD, recession depth.

(RD and RT). RD measurement positively associated with RT agreements (both intra- and inter-); and RT measurement positively associated with RD agreements (only inter-). It seems that more challenging cases for deeper RD or loss of interproximal attachment present higher repeatability of diagnosis, as opposed to recessions with shallower RD and RT1 that offer more ambiguity in the diagnostic process. Finally, RT negatively affected the repeatability of GT at both intra- and inter-operator agreement. RT1 recessions presented a gingival phenotype that was more consistently categorized either as thin or as thick; on the other hand, RT2 and three recessions presented more challenges for the categorization of the gingival phenotype. These findings, valid within the studied population, have to be validated by futures studies using both photographic and chairside measurements.

It is important to highlight that the proper assessment and analysis of GRD characteristics is one of three essential pillars for the effective management of these clinical entities,[8,10] together with a proper identification of the etiologic agent(s) associated with the onset and progression of the defect and the selection of the most appropriate treatment option(s).[8] The use of proper diagnostic and classification tools may reduce the occurrence of errors and the introduction of 'treatment biases' (i.e., errors that lead to a deviation from the true intervention effect) which may lead to a 'downhill' cascade of events (e.g., an inaccurate diagnosis that might lead to an incorrect interpretation of the evidence, inadequate formulation of a treatment plan, and erroneous treatment).[17]

4.2 | Challenges in the determination of the gingival phenotype

Remarkably, evaluation of the gingival phenotype displayed both the "worst" and "best" inter- and intra-examiner agreements. Although KTW may be simply and precisely determined with a linear measurement using a periodontal probe or a caliper, the accurate appraisal of GT can be a challenge in clinical practice.[8] With respect to the low inter- and intra-examiner agreements for GT, this might be explained by the methods used to evaluate this outcome variable. In the present study, GT was assessed in a dichotomous fashion by midfacial transmucosal probe visibility, as shown in Figures 1 and 2. While this assessment method was endorsed in the 2017 World Workshop for its simplicity and relatively high reproducibility (≈80%),[18] based on the findings of this study this notion may be questionable. Previous research has shown that transgingival probing (as known as transmucosal bone sounding) is a simple, precise, and reproducible methods for GT evaluation.[19–21] However, its use is limited due to its invasiveness, which often requires local anesthesia. Direct administration of local anesthesia is also not ideal due to the temporary increase in GT induced by the injected fluid; therefore, application of topical anesthetics is usually recommended. In addition, because the unit measurement of a periodontal probe is 1 mm, precise estimation of values smaller than 1 mm requires the use of other instruments (e.g., endodontic file or reamer with a silicon stopper) instead of a conventional periodontal probe. The measured value could result in an underestimation in case of excessive friction of the silicon stopper on the surface of the instrument during transgingival probing, or overestimation if the silicon stopper is pushed away during the recording. Other methods, such as linear digital imaging evaluation of cone-beam computed tomography in combination with standard tessellation language files, non-ionizing ultrasonography, and optical coherence tomography may be used as reliable alternative methods to measure GT.[8,22] Research is currently ongoing to define the advantages, limitations and cost/benefits for the use of advanced imaging technology in plastic surgery, as well as to understand if these advanced technologies will modify current treatment protocols.[8,22]

4.3 | Development of the 2018 Classification of Gingival Recessions and Gingival Phenotype

The 1985 classification by P.D. Miller[4] has been widely acknowledged as a useful, clinically oriented and reproduceable tool for clinical practice for decades. However, multiple drawbacks related to its application in the context of modern periodontology raised the need for a more comprehensive classification system.[1,10] The possibility of achieving 100% root coverage in teeth with preserved interdental tissue height does not guarantee predictable complete root coverage in all cases of Miller Class I or II. Other important variables such as RD, root surface anomalies, or GT, amongst others, are also neglected. The cutoff between classes in Miller classification is also often unclear and largely subjective. Indeed, Class IV was described as more severe than Class III for interproximal tissue loss and altered tooth positioning, but no tangible cutoff was provided. To overcome these limitations, Cairo et al.[6] regrouped the four Miller classes into three RTs. RT1 merged together Miller Class I and II. RT2 and RT3 better clarified the distinction between Miller Class III and IV, respectively.[6,10] However, the Cairo Classification did not account for important features of the gingival phenotype that were reported in the Miller Classification. The KTW differences that characterized Miller Class I and Class II were not described in Cairo RT1. Also, information on RD, root surface condition or GT were still missing in the Cairo system.[10]

As a result, the 2018 Classification of Gingival Recession Defects and Gingival Phenotype[7,10] was proposed based on Cairo et al. system, with the addition of complementary information related to RD, gingival phenotype (GT and KTW) and tooth condition (RS and CEJ detectability), which are considered standard diagnostic measures nowadays.[7,10] Many short-[23–28] and long-term[29–41] trials have stressed the important role of the gingival phenotype (GT and KTW) and root integrity (RS and CEJ detectability) on the stability of the gingival margin in sites that underwent root coverage therapy. Therefore, it is highly recommended that pretreatment diagnostic assessments adopt the 2018 classification[7,10] or other systems that integrate these features based on available evidence for treatment purposes.[8]

4.4 | Factors affecting reproducibility

The use of classification systems in periodontology stands as a fundamental resource for both clinicians and researchers involved in the diagnosis, treatment and study of periodontal and peri-implant diseases and conditions. An important observation of this study was that, when 16 expert examiners assessed the same sites, a wide variation in their interpretative agreement was observed. This may be largely explained by the method used in this study to evaluate GRD (i.e., evaluation of standardized photographs). The use of photographs allowed international examiners to contribute to data collection without need of traveling for neither examiners or patients. In general, clinical diagnosis should be performed in a clinical setting, in direct contact with the patient, rather than in photographs. However, this clinical chairside component of a study creates barriers related to the need to gather multiple examiners in the same location, at the same time, to evaluate the same patients. This model would not allow the realization of multicenter studies, especially among researchers from international centers. Therefore, for the present research, it was decided to base the diagnosis on photographs as an attempt to test digital technologies, to overcome geographical and temporal barriers created by the on-site examinations, and to make possible international collaboration in a multicenter study on intra- and inter-examiner agreements on the same pool of patients. It might be argued that a direct clinical exam, combined with radiographic/digital imaging may contribute to reduce the amount of inter-examiner variability.[42] Furthermore, studies in other medical disciplines have demonstrated a similar variation in the intra-examiner agreement after examinations at two different timepoints.[43–45] Professional experience, the case scenario and its severity have been suggested as factors that can potentially affect reproducibility.[43] Consequently, the occurrence of small variations in the inter- and intra-examiner assessment can be expected, but these should not alter the core components of the treatment plan. However, a lack of knowledge in properly identifying the type of GRD and its associated characteristics (e.g., gingival phenotype) may lead to under- or over-estimation of individual treatment needs and result in suboptimal therapeutic outcomes.

5 | CONCLUSIONS

Based on our findings and within the discussed limitations, it can be concluded that the different components forming the diagnostic matrix were reproduceable when analyzed individually, and that the 2018 Classification of Gingival Recession Defects and Gingival Phenotype is clinically reproduceable within the examiners. However, the between-examiner agreement for the complete classification matrix was lower. Of the variables assessed, RD and KTW exhibited the highest levels of agreement, likely because the photographs were taken with the presence of a periodontal probe. Conversely, GT assessment via sulcus probe visibility was the least reproduceable assessment.

AUTHOR CONTRIBUTIONS

Giovanpaolo Pini Prato, Leandro Chambrone, and Riccardo Di Gianfilippo conceived and designed the study; Edward P. Allen, Sofia Aroca, Gustavo Avila-Ortiz, Philippe Bouchard, Francesco Cairo, Tali Chackartchi, Pierpaolo Cortellini, Debora Franceschi, Ajay Mahajan, Evelyn A. Mancini, Michael K. McGuire, Tea Natsvlishvili, Mauro P. Santamaria, E. Todd Scheyer, and Hom-Lay Wang examined the cases and collected the data; Claudio M. Pannuti contributed to the design and analyzed the data; Hom-Lay Wang and Kenneth Kornman contributed to data interpretation; and Leandro Chambrone and Giovanpaolo Pini Prato led the writing; All authors critically revised the manuscript.

ACKNOWLEDGMENTS

All the authors declare no conflicts of interest in relation to this study and do not have any financial interests, either directly or indirectly.

"Science brings together what politics tears apart" Tomas Koltai, MD, PhD.

ORCID

Riccardo Di Gianfilippo https://orcid.org/0000-0003-2579-9464
Claudio M. Pannuti https://orcid.org/0000-0003-4181-3975
Francesco Cairo https://orcid.org/0000-0003-3781-1715
Hom-Lay Wang https://orcid.org/0000-0003-4238-1799
Kenneth Kornman https://orcid.org/0000-0003-2252-7324
Leandro Chambrone https://orcid.org/0000-0002-2838-1015

REFERENCES

1. Pini Prato G. The Miller classification of gingival recession: limits and drawbacks. *J Clin Periodontol*. 2011;38:243-245.
2. Sullivan HC, Atkins JH. Free autogenous gingival grafts. 3. Utilization of grafts in the treatment of gingival recession. *Periodontics*. 1968;6:152-160.
3. Sullivan HC, Atkins JH. Free autogenous gingival grafts. I. Principles of successful grafting. *Periodontics*. 1968;6:121-129.
4. Miller PD, Jr. A classification of marginal tissue recession. *Int J Periodontics Restorative Dent*. 1985;5(2):8-13.
5. Pini Prato G, Franceschi D, Cairo F, Nieri M, Rotundo R. Classification of dental surface defects in areas of gingival recession. *J Periodontol*. 2010;81:885-890.
6. Cairo F, Nieri M, Cincinelli S, Mervelt J, Pagliaro U. The interproximal clinical attachment level to classify gingival recessions and predict root coverage outcomes: an explorative and reliability study. *J Clin Periodontol*. 2011;38:661-666.
7. Cortellini P, Bissada NF. Mucogingival conditions in the natural dentition: narrative review, case definitions, and diagnostic considerations. *J Periodontol*. 2018;89 Suppl 1:S204-S213.
8. Chambrone L, Avila-Ortiz G. An evidence-based system for the classification and clinical management of non-proximal gingival recession defects. *J Periodontol*. 2021;92:327-335.
9. Murphy EA. *The Logic of Medicine*. 2nd ed. The Johns Hopkins University Press; 1997:119-136.

10. Pini Prato G, Di Gianfilippo R. On the value of the 2017 classification of phenotype and gingival recessions. *J Periodontol*. 2021;92:613-618.
11. Romandini M, Soldini MC, Montero E, Sanz M. Epidemiology of mid-buccal gingival recessions in nhanes according to the 2018 world workshop classification system. *J Clin Periodontol*. 2020;47:1180-1190.
12. Rotundo R, Nieri M, Mori M, Clauser C, Pini Prato G. Aesthetic perception after root coverage procedure. *J Clin Periodontol*. 2008;35:705-712.
13. Rotundo R, Nieri M, Bonaccini D, et al. The smile esthetic index (SEI): a method to measure the esthetics of the smile. An intra-rater and inter-rater agreement study. *Eur J Oral Implantol*. 2015;8:397-403.
14. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med*. 1998;17:101-110.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174
16. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155-163.
17. Chambrone L, Garcia-Valenzuela FS, Avila-Ortiz G. Errors and complications in clinical periodontal practice due to methodologic bias and bad interpretation of the evidence. *Periodontol 2000*. 2022;Epub ahead of print.
18. De Rouck T, Eghbali R, Collys K, De Bruyn H, Cosyn J. The gingival biotype revisited: transparency of the periodontal probe through the gingival margin as a method to discriminate thin from thick gingiva. *J Clin Periodontol*. 2009;36:428-433.
19. Frost NA, Mealey BL, Jones AA, Huynh-Ba G. Periodontal biotype: gingival thickness as it relates to probe visibility and buccal plate thickness. *J Periodontol* 2015;86:1141-1149.
20. Hutton CG, Johnson GK, Barwacz CA, Allareddy V, Avila-Ortiz G. Comparison of two different surgical approaches to increase peri-implant mucosal thickness: a randomized controlled clinical trial. *J Periodontol*. 2018;89:807-814.
21. Sala L, Alonso-Perez R, Agustin-Panadero R, Ferreiroa A, Carrillo-de-Albornoz A. Comparative in vitro study of two methods for gingival biotype assessment. *J Clin Exp Dent*. 2018;10:e858-e863.
22. Chambrone L, Avila-Ortiz G. The tissues. In: Chambrone L, Avila-Ortiz G, eds. *Tissues: Critical Issues in Periodontal and Implant-Related Plastic and Reconstructive Surgery*. Quintessence Publishing; 2022:54-55.
23. Chambrone L, Tatakis DN. Periodontal soft tissue root coverage procedures: a systematic review from the AAP Regeneration Workshop. *J Periodontol* 2015;86(2 Suppl):S8-S51.
24. Chambrone L, Ortega MAS, Sukekava F, et al. Root coverage procedures for treating single and multiple recession-type defects: An updated Cochrane systematic review. *J Periodontol* 2019;90:1399-1422.
25. Chambrone L, Boltelho J, Machado V, Mascarenhas P, Mendes JJ, Avila-Ortiz G. Does the subepithelial connective tissue graft in conjunction with a coronally advanced flap remain as the gold standard therapy for the treatment of single gingival recession defects? A systematic review and network meta-analysis. *J Periodontol*. 2022;93:1336-1352.
26. Chambrone L, Barootchi S, Avila-Ortiz G. Efficacy of biologics in periodontal plastic surgery: an American academy of periodontology best evidence systematic review and network meta-analysis. *J Periodontol*. 2022. https://doi.org/10.1002/JPER.22-0075. Epub ahead of print.
27. Baldi C, Pini-Prato G, Pagliaro U, et al. Coronally advanced flap procedure for root coverage. Is flap thickness a relevant predictor to achieve root coverage? A 19-case series. *J Periodontol*. 1999;70:1077-1084.
28. Cairo F, Cortellini P, Pilloni A, et al. Clinical efficacy of coronally advanced flap with or without connective tissue graft for the treatment of multiple adjacent gingival recessions in the aesthetic area: a randomized controlled clinical trial. *J Clin Periodontol*. 2016;43:849-856.
29. Chambrone L, Tatakis DN. Long-term outcomes of untreated buccal gingival recessions: a systematic review and meta-analysis. *J Periodontol*. 2016;87:796-808.
30. Pini Prato GP, Magnani C, Chambrone L. Long-term evaluation (20 years) of the outcomes of coronally advanced flap in the treatment of single recession-type defects. *J Periodontol*. 2018;89:265-274.
31. Pini Prato GP, Franceschi D, Cortellini P, Chambrone L. Long-term evaluation (20 years) of the outcomes of subepithelial connective tissue graft plus coronally advanced flap in the treatment of maxillary single recession-type defects. *J Periodontol*. 2018;89:1290-1299.
32. Agudio G, Chambrone L, Selvaggi F, Pini-Prato GP. Effect of gingival augmentation procedure (free gingival graft) on reducing the risk of non-carious cervical lesions: a 25- to 30-year follow-up study. *J Periodontol*. 2019;90:1235-1243.
33. Barootchi S, Tavelli L, Di Gianfilippo R, et al. Soft tissue phenotype modification predicts gingival margin long-term (10-year) stability: longitudinal analysis of six randomized clinical trials. *J Clin Periodontol*. 2022;49:672-683. https://doi.org/10.1111/jcpe.13641. Epub 2022 May 13.
34. Franceschi D, Pini Prato GP, Di Gianfilippo R Double connective tissue graft to treat deep coronal-radicular abrasion: a 19-year follow-up case report. *Clin Adv Periodontics*. 2021;11:171-175. https://doi.org/10.1002/cap.10176. Epub 2021 Jul 20.
35. Di Gianfilippo R, Wang IC, Steigmann L, Velasquez D, Wang HL, Chan HL. Efficacy of microsurgery and comparison to macrosurgery for gingival recession treatment: A systematic review with meta-analysis. *Clin Oral Investig*. 2021;25:4269-4280. Epub 2021 Apr 29.
36. Pini Prato GP, Mancini EA, Di Gianfilippo R, Franceschi D. Periodontal plastic approaches for reshaping the mucogingival junction following grafting procedures in patients requesting esthetic improvement. *Int J Periodontics Restorative Dent*. 2021;41:207-214.
37. Tavelli L, Barootchi S, Di Gianfilippo R, et al. Patient experience of autogenous soft tissue grafting has an implication for future treatment: a 10-15-year cross-sectional study. *J Periodontol*. 2021;92:637-647.
38. Barootchi S, Tavelli L, Di Gianfilippo R, et al. Gingival phenotype modification as a result of root coverage procedure with two human dermal matrices: long-term assessment of a randomized clinical trial. *Int J Periodontics Restorative Dent*. 2021;41:719-726.
39. Barootchi S, Tavelli L, Di Gianfilippo R, et al. Acellular dermal matrix for root coverage procedures. 9-year assessment of treated

isolated gingival recessions and their adjacent untreated sites. *J Periodontol*. 2021;92:254-262.
40. Barootchi S, Tavelli L, Di Gianfilippo R, et al. Long term assessment of root coverage stability using connective tissue graft with or without an epithelial collar for gingival recession treatment. a 12-year follow-up from a randomized clinical trial. *J Clin Periodontol*. 2019;46:1124-1133.
41. Tavelli L, Barootchi S, Di Gianfilippo R, et al. Acellular dermal matrix and coronally advanced flap or tunnel technique in the treatment of multiple adjacent gingival recessions. a 12-year follow-up from a randomized clinical trial. *J Clin Periodontol*. 2019;46:937-948.
42. Nieri M, Giuntini V, Pagliaro U, Giani M, Franchi L, Franceschi D. Efficacy of a u-shaped automatic electric toothbrush in dental plaque removal: a cross-over randomized controlled trial. *Int J Environ Res Public Health*. 2020;17:4649.
43. Jackson SL, Frederick PD, Pepe MS, et al. Diagnostic reproducibility: what happens when the same pathologist interprets the same breast biopsy specimen at two points in time? *Ann Surg Oncol*. 2017;24:1234-1241.
44. Elmore JG, Longton GM, Carney PA, et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *JAMA*. 2015;313:1122-1132.
45. Brunyé TT, Mercan E, Weaver DL, Elmore JG. Accuracy is in the eyes of the pathologist: the visual interpretive process and diagnostic accuracy with digital whole slide images. *J Biomed Inform*. 2017;66:171-179.

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: 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*. 2022;1-12. https://doi.org/10.1002/JPER.22-0501