Received: 11 March 2024 | Revised: 1 July 2024 | Accepted: 30 July 2024

DOI: 10.1002/JPER.24-0173

JOURNAL OF Periodontology

CASE SERIES

Diagnostic reproducibility of the 2018 Classification of Gingival Recessions: Comparing photographic and in-person diagnoses

Riccardo Di Gianfilippo[1,2,3] | **GiovanPaolo Pini Prato**[4] | **Debora Franceschi**[5] | **Walter Castelluzzo**[6] | **Luigi Barbato**[6] | **Alessandra Bandel**[6] | **Maria Di Martino**[6] | **Claudio M. Pannuti**[7] | **Leandro Chambrone**[8,9,10] | **Francesco Cairo**[11]

[1]Department of Periodontics & Oral Medicine, University of Michigan School of Dentistry, Ann Arbor, Michigan, USA

[2]Pacific Academy of Periodontal and Implant Research, Bellevue, Washington, USA

[3]Foundation for Oral Rehabilitation, Lucerne, Switzerland

[4]Tuscany Academy of Dental Research, Florence, Italy

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

[6]Unit in Periodontology and Periodontal Medicine, The University of Florence, Florence, Italy

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

[8]Evidence-Based Hub, Centro de Investigação Interdisciplinar Egas Moniz, Egas Moniz-Cooperativa de Ensino Superior, Caparica, Portugal

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

[10]Unit of Basic Oral Investigation, Universidad El Bosque, Bogota, Colombia

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

Correspondence
Riccardo Di Gianfilippo, Department of Periodontics & Oral Medicine, University of Michigan School of Dentistry, Ann Arbor, MI 48104, USA.
Email: rdgianfi@umich.edu

Abstract

Background: To assess how the diagnostic reproducibility of the 2018 Classification of Gingival Recession Defects (GRD) could be applied when comparing in-person chairside measurements with photographic measurements.

Methods: Thirty-four GRD were photographed and evaluated by 4 masked operators. For each case, the operators measured twice 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 (RSs), chairside, and on photographs. Intraclass correlation coefficient (ICC) with 95% confidence intervals (CI) was calculated 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: RD, KTW, and RT showed excellent overall intra-operator agreement (> 0.93), and from good to excellent overall inter-operator agreement (> 0.80), for both clinical and photographic measurements. Agreements were lower for GT,

CEJ, and RS. Overall clinical and photographic agreements were within 0.1 difference for every variable, except for inter-operator agreement for RS which was 0.72 for clinical measurements and 0.45 for photographic measurements. The lowest overall agreement between clinical versus photographic measurements existed for CEJ (0.28) and RS (0.35).

Conclusions: Variables composing the 2018 Classification of GRD are reproducible, both clinically and on photographs, with comparable agreements. The overall agreement was higher for KTW, RD, and RT, and lower for GT, CEJ, and RS, for both clinical and photographic measurements. The comparison between chairside and photographic evaluations indicated fair to excellent agreement for most variables, with CEJ and RS showing fair agreement.

KEYWORDS
classification, gingival recession, periodontics, phenotype, tooth root

Plain Language Summary

As digital diagnostics evolve to facilitate clinical diagnostic measurement, we aimed to assess the effectiveness of intraoral photography for diagnosing gingival recession defects (GRD) according to the 2018 Classification of GRD, compared to traditional clinical examination.

Standardized photographs of thirty-four GRD cases were captured. Four masked operators evaluated the same gingival recessions twice in a clinical setting and twice using photographs. Measurement repeatability within and between operators was calculated for both clinical and photographic settings, and the two settings were compared.

Continuous measurements such as recession depth and keratinized tissue width, as well as the evaluation of interproximal attachment height (recession type), showed excellent agreement both clinically and photographically. Agreement was lower for gingival thickness and the detectability of tooth anatomical landmarks, such as the cemento-enamel junction and the presence of root steps. Overall, the agreement between chairside and photographic evaluations was generally good, but lower when evaluating tooth anatomical landmarks.

The variables composing the 2018 Classification of GRD are reproducible in both clinical and photographic settings, with comparable levels of agreement. However, there was consistently worse agreement for gingival thickness and when evaluating tooth anatomical landmarks.

1 | INTRODUCTION

Classification systems are essential tools for clinicians and researchers in the medical and dental fields, including periodontology. Accurate classifications are pivotal, as they form the basis for proper diagnosis. This, in turn, is essential for effective treatment and helps prevent potential errors in therapeutic approaches.[1–3] The precise diagnosis of the features associated with gingival recession defects (GRD) is of paramount value within the field of periodontal plastic surgery. It aids in determining the severity of the condition, indications for therapy, predicting treatment outcomes, and evaluating the results over short- and long-term periods.[4–9]

Historically, several classifications have been adopted to describe GRD.[10–13] Each of these classifications has reflected the state-of-the-art clinical practices and knowledge of its time.[14] A notable milestone occurred during the 2017 World Workshop on Periodontal and Peri-implant Diseases and Conditions when a new classification for gingival recessions was proposed,[15] commonly referred to as the 2018 Classification of GRD and Gingival Phenotype. This innovative classification uses a 4 × 5 matrix, integrating the most relevant clinical variables from previous systems. It includes midfacial recession depth (RD), interproximal recession type (RT), gingival thickness (GT), keratinized tissue width (KTW), detectability of the cemento–enamel junction (CEJ), and the presence of non-carious cervical lesions or root steps (RSs).[14,15] The 2018 Classification of GRD is considered the most current and comprehensive diagnostic system available for assessing gingival recession. It stands out because it can simultaneously evaluate recession-related features, gingival phenotypic characteristics, and root surface attributes.[14]

Since its publication, various studies have examined the application of this classification for clinical and research purposes, including its use in epidemiological studies and clinical trials on gingival recessions.[9,14,16,17] A recent multicenter study evaluated the diagnostic reproducibility of the 2018 Classification of GRD and Gingival Phenotype among international experts.[18] The classification was found reproducible within and between examiners when the clinical variables were analyzed individually. The highest agreement was observed in continuous measurements such as KTW and RD, while GT showed the least agreement.

Notably, the methodology employed in the evaluation was exclusively based on standardized photographs.[18] Clinical diagnosis is typically performed in a clinical setting with patients. However, the use of photographs was chosen to overcome logistical challenges of bringing multiple examiners in a single location together at the same time, which is required for clinical chairside evaluations. This approach was crucial for a study with operators from various international centers.[18] The study opened the field of periodontal plastic surgery for application in telemedicine, allowing remote operators to provide a diagnosis of gingival recessions located in geographically far centers.

It might be argued that direct clinical examinations could have reduced the inter-examiner variability compared to the photographic diagnosis for the 2018 Classification of GRD and gingival phenotype in the previous study.[18] However, this hypothesis remains to be tested. Therefore, the objective of the present study was to assess the diagnostic reproducibility of the 2018 Classification of GRD, by estimating the agreement between the in-person chairside evaluation and photographic evaluation.

2 | MATERIALS AND METHODS

The present human study was conducted in accordance with the Helsinki Declaration as revised in 2013. The study protocol was reviewed and approved by the local Ethical Review Board (ID: 23979 oss. by CEAVC Comitato Etico Regionale per Sperimentazione Clinica della Regione Toscana). This study was designed according to the STARD 2015 Guidelines for Reporting Diagnostic Accuracy Studies.[19]

A total of 34 single GRD from 34 consecutive subjects were screened from patients attending the Unit of Periodontology and Periodontal Medicine of the University of Florence (Italy), and enrolled for this reproducibility study. All patients provided written informed consent. The sample consisted of 21 maxillary and 13 mandibular buccal GRD counting 8 incisors, 8 canines, and 18 premolars.

2.1 | Clinical activities

Before starting the study, 4 trained operators (D.F., W.C., L.B., F.C.) underwent a calibration session on the variables featuring the 2018 Classification of GRD and gingival phenotype. Definitions of RT, RD, GT, KTW, CEJ detectability, and RS were reviewed and agreed upon by all examiners. Subsequently, all operators examined 2 patients exclusively invited for calibration purposes, who were not included in the later phases of the study. The calibration session was deemed successful when a perfect match of values for RT, RD, GT, KTW, CEJ, and RS was achieved among all examiners. Following this, the operators commenced the photographic and clinical activities on the final population pool.

Clinical activities started with the photographic documentation of sites with GRD, which was performed to collect a series of 7 standardized intraoral images in frontal and lateral views, with and without a UNC-15 periodontal probe[*] (Figure 1). All photos were taken using the same photographic settings and equipment.[†,‡,§] Each image underwent a rigorous quality-control process immediately after acquisition, while the patient remained in the dental chair. Quality control involved ensuring a perpendicular view of the involved tooth for frontal images, verifying the correct positioning of the probe lining against the tooth surface, and ensuring realistic color visibility. If necessary, images were reacquired during the same clinical session.

[*] CP-15 UNC periodontal probe, Hu-Friedy, Chicago, Illinois.
[†] Nikon D3400 camera body, Nikon, Minato City, Tokyo, Japan.
[‡] Flash Nissin Mf18, Nissin Digital, Ukyo-ku, Kyoto, Japan.
[§] Nikon D3400 camera body, Nikon, Minato City, Tokyo, Japan.

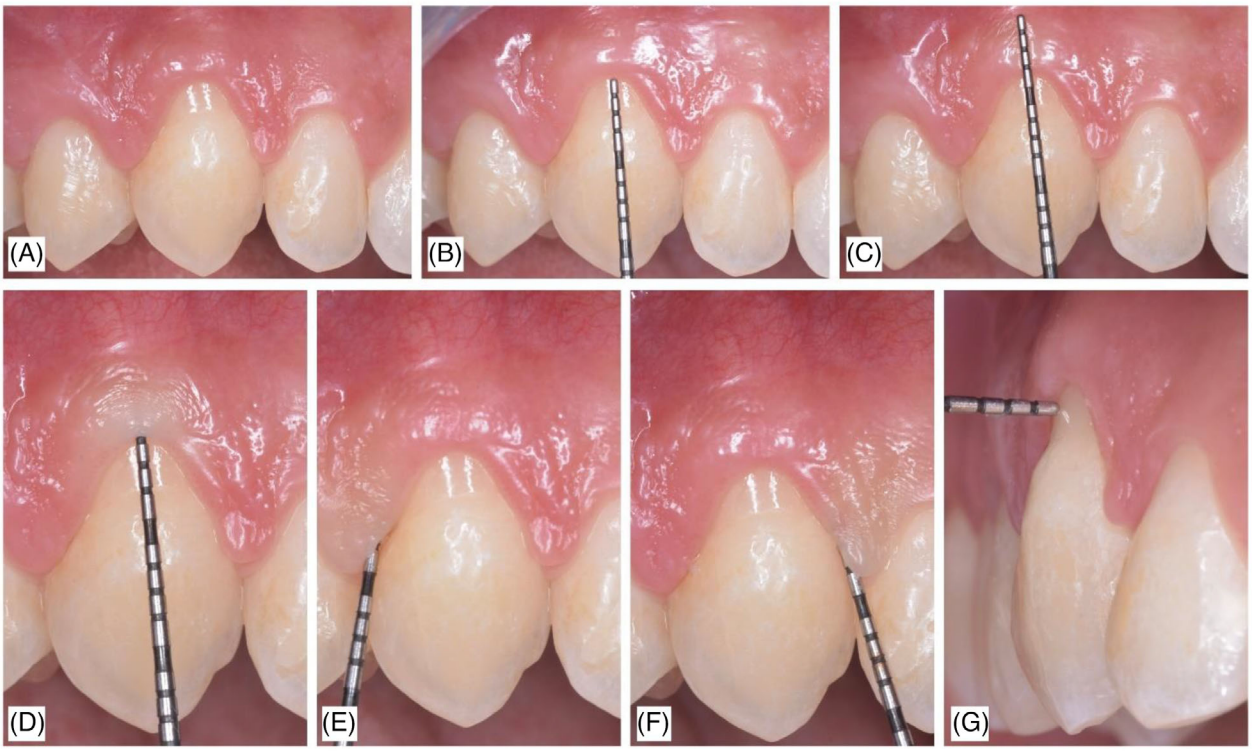

FIGURE 1 Photographic documentation of a maxillary right canine illustrated with a series of standardized intraoral images from a frontal and lateral view, with and without a UNC-15 probe. Frontal image was captured to provide an overview of the gingival recession defect (A), recession depth (B), keratinized tissue width (C), gingival thickness (D), distal (E), and mesial (F) interproximal probing were used for calculation of the recession type. (G) Lateral view of the gingival recession defect allowed measurement of the depth of the non-carious cervical lesion. The integrity of the cemento–enamel junction was assessed at frontal and lateral view.

After completion of the sets of pictures, the 4 operators (D.F., W.C., F.C., L.B.) evaluated the GRD clinically based on the features of the 2018 Classification of GRD and gingival phenotype. The variables measured for each recession were as follows: RT 1, 2, or 3; RD in mm; GT (thin vs. thick) evaluated by tissue transparency upon midfacial sulcus probing; KTW in mm; integrity of the CEJ (A vs. B); and presence or absence of RS deeper than 1 mm in the horizontal dimension (+ vs. −). The examiners were instructed to round the measurements to the nearest 1.0 mm for RD and KTW.

During the clinical measurements on the patient, each operator measured each variable twice for each recession. The 2 measurements by the same operator on the same recession were not consecutive. Operators were blinded to both their own previous clinical measurements and any measurements obtained by other operators. By the end of the clinical session, each recession received 8 measurements for each variable from 4 operators.

Two weeks after completing the clinical activities, the operators repeated the diagnostic measurements on the images of the recessions as reported in Pini Prato et al.[18] The examiners were instructed to assess each recession twice, with a 2-week gap between the first and second measurement cycles. They were taught to round the measurements to the nearest 1.0 mm for RD and KTW. They were also instructed not to reference their prior assessments during the second evaluation.

2.2 | Statistical analysis

The sample size for this study was determined using the method described by Walter et al.[20] (i.e., agreement assessment by the calculation of intraclass correlation coefficient (ICC) [p]). Taking into consideration 4 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 29 GRD were required. Taking into account a 20% increase in the minimum sample size, the final sample size for this study was determined to be 34 GRD.

Intra- and inter-examiner agreements were calculated for the clinical and photographic measurements. ICC (2-way random, average agreement) with 95% confidence intervals (95% CI) were used for continuous variables (RD and KTW). Kappa with 95% CI was used to assess agreement for qualitative nominal variables (GT, CEJ, RS).

For the ordinal variable RT, quadratic weighted Kappa with 95% CI was used. In the case of the overall inter-examiner agreement, we considered examiner #3 as the gold standard.

Kappa values were interpreted using: poor agreement (< 0.00), slight agreement (0.00–0.20), fair agreement (0.21–0.40), moderate agreement (0.41–0.60), substantial agreement (0.61–0.80), and almost perfect agreement (0.81–1.00), according to Landis and Koch.[21] ICC values were categorized as poor agreement (< 0.5), moderate agreement (0.5–0.75), good agreement (0.75–0.9), and excellent agreement (> 0.90) based on the criteria from Koo and Li.[22]

The agreement between the clinical and photographic measurements was calculated using the ICC with 95% CI for continuous variables (RD and KTW). For qualitative variables (GT, CEJ, RS), we used Kappa with 95% CI to assess agreement. For the ordinal variable RT, we used the quadratic weighted Kappa. Inter- and intra-examiner agreement was calculated in JAMOVI version 1.6.23 using the SimplyAgree and MedDecide modules.

The effect of predictor variables (RD, RT, KTW, RS, CEJ, and GT) on the intra-operator agreements for both photographic and clinical measurements (Agree = 1, Disagree = 0) was calculated using 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. Analysis of variance (ANOVA) with Wald test was conducted to verify the significance of the predictors (independent variables). The standard deviation coefficients were reported, as well as the odds ratio (OR), which indicates an increase or decrease in the chance of agreement, with the corresponding CI. The adjustment was made by the R program, using the glm function, from the "stats" package. Then, model selection was done using the step-forward method with the "step" function from the same package, in which 1 predictor variable is inserted at a time and tested using the Bayesian Information Criterion. Variables that reduced the Akaike Information Criterion (AIC) were retained in the model. The significance level of statistical tests was set at 5%.

3 | RESULTS

3.1 | Agreement for clinical measurements

Intra-examiner agreement of RD and KTW was considered excellent, as ICC ranged from 0.96 to 1.00 for RD, and from 0.96 to 0.98 for KTW, considering each of the 4 different examiners. Intra-examiner agreement of GT, CEJ, and RS ranged from substantial (Kappa = 0.70) to almost perfect (Kappa = 0.93) for GT; from moderate (Kappa = 0.60) to almost perfect (Kappa = 1.00) for CEJ, and it was almost perfect (Kappa = 0.85–1.00) for RS, considering the 4 different examiners. Intra-examiner assessment for RT showed an almost perfect (weighted Kappa 0.94–1.00) agreement among the 4 different examiners (Table 1). Overall intra-examiner agreement was considered excellent for RD (ICC = 0.98) and KTW (ICC = 0.97). The agreement was almost perfect for GT (Kappa = 0.84), CEJ (Kappa = 0.88), RS (Kappa = 0.93), and RT (weighted kappa = 0.97) (Table 1).

Inter-examiner agreement ranged from 0.89 to 0.95 for RD, and from 0.84 to 0.91 for KTW, when considering the agreement between all examiners. Inter-examiner agreement of GT, CEJ, and RS was calculated with Kappa and ranged from 0.35 to 0.80 (GT), 0.24 to 0.63 (CEJ), and 0.40 to 0.90 (RS) between all examiners. Weighted kappa was used to assess inter-examiner agreement for RT, and ranged from 0.88 to 0.97 (Table 1). When the gold-standard operator (#3) was compared with all other operators, overall inter-examiner agreement was excellent for RD (ICC = 0.94), and it was good for KTW (ICC = 0.88). Overall inter-examiner agreement was moderate for GT (kappa = 0.41) and for CEJ (kappa = 0.49), and it was substantial for RS (kappa = 0.72) and RT (weighted kappa = 0.80) (Table 1).

For both ordinal (RT) and continuous (RD and KTW) variables, the disagreement between measurements occurred, at most, by 1 unit, configuring just 1 level of disagreement. In other words, for RT, the disagreement was at most of 1 score of difference, and for the continuous variables, the disagreement was, at most, of 1 mm.

3.2 | Agreement for photographic measurements

Intra-examiner agreement of RD and KTW was excellent, with ICC varying between 0.96 and 1.00 for RD, and between 0.94 and 1.00 for KTW across each of the 4 different examiners. The intra-examiner agreement for GT varied from substantial (Kappa = 0.63) to almost perfect (Kappa = 0.93); for CEJ it ranged from substantial (Kappa = 0.67) to almost perfect (Kappa = 1.00), and for RS it was almost perfect (Kappa = 0.82–1.00), considering the 4 different examiners. Intra-examiner agreement for RT was almost perfect (weighted Kappa 0.83–1.00) across all 4 different examiners (Table 2). Overall intra-examiner agreement was excellent for RD (ICC = 0.96) and KTW (ICC = 0.97). The agreement was substantial for GT (Kappa = 0.80), almost perfect for CEJ (Kappa = 0.84),

TABLE 1 Intra- and inter-examiner agreements for clinical measurements

Intra-examiner agreement for clinical measurements						
Operator	**RD**	**KTW**	**GT**	**CEJ**	**RS**	**RT**
1	0.96 (0.92–0.97)	0.97 (0.94–0.98)	0.93 (0.78–1.00)	0.60 (0.29–0.92)	0.85 (0.65–1.00)	0.97 (0.89–1.00)
2	0.98 (0.97–0.99)	0.98 (0.96–0.99)	0.70 (0.45–0.74)	1.00 (1.00–1.00)	0.90 (0.72–1.00)	0.94 (0.80–1.00)
3	1.00 (1.00–1.00)	0.96 (0.93–0.97)	0.88 (0.72–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
4	0.96 (0.93–0.98)	0.96 (0.93–0.98)	0.87 (0.71–1.00)	0.86 (0.67–1.00)	1.00 (1.00–1.00)	0.97 (0.91–1.00)
Overall (all operators)	0.98 (0.97–0.98)	0.97 (0.96–0.98)	0.84 (0.75–0.93)	0.88 (0.79–0.96)	0.93 (0.86–1.00)	0.97 (0.93–1.00)
Inter-examiner agreement for clinical measurements						
Operator	**RD**	**KTW**	**GT**	**CEJ**	**RS**	**RT**
1 versus 2	0.90 (0.83–0.94)	0.87 (0.76–0.92)	0.49 (0.18–0.80)	0.34 (0.00–0.71)	0.46 (0.07–0.83)	0.88 (0.72–1.00)
1 versus 3	0.93 (0.87–0.96)	0.88 (0.79–0.93)	0.35 (0.03–0.66)	0.33 (0.00–0.68)	0.76 (0.71–1.00)	0.97 (0.89–1.00)
1 versus 4	0.94 (0.89–0.96)	0.84 (0.73–0.91)	0.80 (0.59–1.00)	0.24 (0.00–0.63)	0.40 (0.01–0.79)	0.79 (0.60–0.97)
2 versus 3	0.93 (0.88–0.96)	0.91 (0.84–0.95)	0.47 (0.17–0.76)	0.54 (0.24–0.84)	0.74 (0.47–1.00)	0.91 (0.78–1.00)
2 versus 4	0.89 (0.81–0.94)	0.86 (0.76–0.92)	0.69 (0.44–0.94)	0.63 (0.33–0.93)	0.90 (0.72–1.00)	0.74 (0.55–0.94)
3 versus 4	0.95 (0.91–0.97)	0.84 (0.71–0.91)	0.41 (0.00–0.71)	0.60 (0.31–0.89)	0.67 (0.37–0.97)	0.77 (0.59–0.95)
Overall (#3 versus other operators)	0.94 (0.91–0.95)	0.88 (0.83–0.91)	0.41 (0.25–0.58)	0.49 (0.31–0.67)	0.72 (0.56–0.88)	0.80 (0.70–0.90)

Note: RD, KTW, GT, detectability of the CEJ, RS, and RT are reported by the operator. The overall inter-examiner agreement was calculated as the gold standard (operator #3) versus other operators. Values are presented as ICC (95% CI) for RD and KTW, Kappa (95% CI) for GT, CEJ, and RS, and quadratic weighted Kappa (95% CI) for RT.

Abbreviations: CEJ, cemento–enamel junction; CI, confidence interval; GT, gingival thickness; ICC, intraclass correlation coefficient; KTW, keratinized tissue width; RD, recession depth; RS, root step; RT, recession type.

TABLE 2 Intra- and inter-examiner agreements for photographic measurements

Intra-examiner agreement for photographic measurements						
Operator	**RD**	**KTW**	**GT**	**CEJ**	**RS**	**RT**
1	0.98 (0.97–0.99)	1.00 (1.00–1.00)	0.93 (0.79–1.00)	0.79 (0.52–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
2	0.96 (0.94–0.98)	0.95 (0.92–0.97)	0.63 (0.35–0.89)	0.67 (0.32–1.00)	1.00 (1.00–1.00)	0.92 (0.80–1.00)
3	0.89 (0.79–0.94)	0.94 (0.90–0.96)	0.75 (0.52–0.98)	0.75 (0.53–0.98)	0.82 (0.58–1.00)	0.83 (0.63–1.00)
4	1.00 (1.00–1.00)	0.99 (0.99–0.99)	0.93 (0.79–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.97 (0.92–1.00)
Overall (all operators)	0.96 (0.95–0.97)	0.97 (0.96–0.98)	0.80 (0.69–0.91)	0.84 (0.73–0.94)	0.93 (0.84–1.00)	0.93 (0.87–0.99)
Inter-examiner agreement for photographic measurements						
Operator	**RD**	**KTW**	**GT**	**CEJ**	**RS**	**RT**
1 versus 2	0.95 (0.92–0.97)	0.96 (0.89–0.96)	0.80 (0.59–1.00)	0.89 (0.68–1.00)	1.00 (1.00–1.00)	0.91 (0.78–1.00)
1 versus 3	0.88 (0.72–0.94)	0.93 (0.81–0.96)	0.28 (0.00–0.63)	0.37 (0.02–0.72)	0.47 (0.05–0.90)	0.84 (0.66–1.00)
1 versus 4	0.96 (0.93–0.98)	0.94 (0.86–0.97)	0.57 (0.26–0.88)	0.09 (0.00–0.48)	0.52 (0.01–1.00)	0.87 (0.71–1.00)
2 versus 3	0.92 (0.81–0.96)	0.92 (0.82–0.96)	0.25 (0.00–0.59)	0.29 (0.00–0.66)	0.47 (0.05–0.90)	0.87 (0.71–1.00)
2 versus 4	0.96 (0.94–0.98)	0.92 (0.85–0.96)	0.54 (0.24–0.84)	0.15 (0.00–0.53)	0.52 (0.01–1.00)	0.79 (0.62–0.97)
3 versus 4	0.93 (0.87–0.96)	0.94 (0.91–0.97)	0.41 (0.08–0.74)	0.62 (0.35–0.89)	0.40 (0.05–0.75)	0.78 (0.60–0.95)
Overall (#3 versus other operators)	0.91 (0.83–0.95)	0.93 (0.89–0.95)	0.31 (0.11–0.51)	0.44 (0.24–0.63)	0.45 (0.20–0.69)	0.82 (0.73–0.92)

Note: RD, KTW, GT, detectability of the CEJ, RS, and RT are reported by the operator. The overall inter-examiner agreement was calculated as the gold standard (operator #3) versus other operators. Values are presented as ICC (95% CI) for RD and KTW, Kappa (95% CI) for GT, CEJ, and RS, and quadratic weighted Kappa (95% CI) for RT.

Abbreviations: CEJ, cemento–enamel junction; CI, confidence interval; GT, gingival thickness; ICC, intraclass correlation coefficient; KTW, keratinized tissue width; RD, recession depth; RS, root step; RT, recession type.

almost perfect for RS (Kappa = 0.93), and almost perfect for RT (weighted kappa = 0.93) (Table 2).

Inter-examiner agreement of RD and KTW ranged from 0.88 to 0.96 for RD, and from 0.92 to 0.96 for KTW, when considering the agreement between all examiners. Inter-examiner agreement of GT, CEJ, and RS was calculated with Kappa and ranged from 0.25 to 0.80 (GT), 0.09 to 0.89 (CEJ), and 0.40 to 1.00 (RS) between all examiners. Weighted kappa was used to assess inter-examiner agreement for RT, and ranged from 0.78 to 0.91 (Table 2). When the gold-standard operator (#3) was compared with all other operators, overall inter-examiner agreement was excellent for RD (ICC = 0.91) and KTW (ICC = 0.93). Overall Inter-examiner agreement was fair for GT (kappa = 0.31), and moderate for CEJ (kappa = 0.44) and RS (kappa = 0.45). An almost perfect agreement was observed for RT (weighted kappa = 0.82) (Table 2).

For both ordinal (RT) and continuous (RD and KTW) variables, the disagreement between measurements occurred, at most, by 1 unit, configuring just 1 level of disagreement. In other words, for RT, the disagreement was at most of 1 score of difference, and for the continuous variables, the disagreement was, at most, of 1 mm.

3.3 | Effect of individual variables on the agreement

A significant negative effect was observed for RT on the clinical intra-operator agreement of RD (OR = 0.24, $p < 0.001$). For KTW, there was a significant positive effect on the intra-operator agreement of RD (OR = 1.95, $p < 0.05$), CEJ (OR = 2.32, $p < 0.05$), and RS (OR = 5.33, $p < 0.05$) (see Table S1 in online *Journal of Periodontology*).

With regard to the photographic intra-operator agreement, there was a significant positive effect for RT on the agreement of KTW (OR = 4.4, $p < 0.001$); and for CEJ on the intra-operator agreement of RT (OR = 9.06, $p < 0.05$) (see Table S2 in online *Journal of Periodontology*).

The variance inflation factor indicated multicolinearity for RD. Consequently, it was not included in the final model.

3.4 | Agreement between clinical and photographic measurements

Upon evaluating the agreement between clinical and photographic measurements of RD across the 4 distinct examiners, the results indicated a range from good (ICC = 0.82) to excellent agreement (ICC = 0.93). In contrast, when assessing KTW, the concordance between the 2 methods of measurement demonstrated a moderate to good agreement, with ICC values varying between 0.66 and 0.82. For GT, the agreement ranged from moderate (kappa = 0.41) to substantial (kappa = 0.80), while for CEJ, it ranged from slight (kappa = 0.01) to moderate (kappa = 0.42). For RS, the agreement ranged from slight (kappa = 0.13) to moderate (kappa = 0.47). Agreement between the 2 methods of measuring RT ranged from substantial (kappa = 0.74) to almost perfect (kappa = 0.88) (Table 3).

When all examiners were considered (overall agreement), there was an excellent agreement between clinical and photographic measurements of RD (ICC = 0.92), and a good agreement for KTW (ICC = 0.77). The kappa measure of agreement between clinical and photographic measurements of GT was equal to 0.62, which is considered a substantial agreement. On the other hand, overall agreement between the 2 assessments was 0.28 for CEJ and 0.35 for RS, which can be interpreted as a fair agreement. The overall agreement between clinical and photographic measurements was almost perfect for RT (weighted kappa = 0.82) (Table 3).

4 | DISCUSSION

4.1 | Photographic and clinical agreements

As our field progresses toward a future defined by collaborative efforts between human operators and digital technology aimed at improving human accuracy, conventional diagnostics are being challenged by a new generation of digital diagnostics under evaluation. To our knowledge, this study is the first to compare in-person chairside measurements and photographic measurements to assess the diagnostic reproducibility of the 2018 Classification of GRD.[15]

Both clinical and photographic examinations presented high intra-operator agreement for most variables. Clinical inter-operator agreement was high for continuous variables (RD and KTW), with disagreement limited to 1 mm at most. Clinical inter-operator agreement was lower for categorical variables, with higher agreement for RT (0.80) and lower for GT and CEJ. However, the disagreement for RT remained within the range of 1 category at most. Similarly, for photographic variables, inter-operator agreement was high for continuous variables (RD and KTW), with disagreement limited to 1 mm at most. The agreement was lower for categorical variables, with higher agreement for RT (0.82) and lower for GT, RS, and CEJ. Yet, the disagreement for RT remained within the range of 1 category at most.

Photographic intra-operator agreements were very similar and within the same level of agreement as the clinical

TABLE 3 Agreement between clinical and photographic measurements

Operator	RD	KTW	GT	CEJ	RS	RT
1	0.93 (0.89–0.96)	0.79 (0.63–0.88)	0.72 (0.45–0.97)	0.40 (0.02–0.79)	0.47 (0.05–0.90)	0.88 (0.72–1.00)
2	0.93 (0.88–0.96)	0.82 (0.59–0.91)	0.57 (0.28–0.85)	0.42 (0.03–0.80)	0.62 (0.21–1.00)	0.86 (0.69–1.00)
3	0.82 (0.76–0.93)	0.82 (0.54–0.91)	0.41 (0.10–0.71)	0.01 (0.00–0.35)	0.13 (0.00–0.55)	0.82 (0.64–0.99)
4	0.92 (0.84–0.96)	0.66 (0.37–0.81)	0.80 (0.59–1.00)	0.33 (0.00–0.68)	0.25 (0.00–0.74)	0.74 (0.57–0.90)
Overall (all operators)	0.92 (0.89–0.94)	0.77 (0.64–0.85)	0.62 (0.48–0.75)	0.28 (0.09–0.46)	0.35 (0.13–0.57)	0.82 (0.73–0.91)

Note: RD, KTW, GT, CEJ detectability, RS, and RT by the operator. Values are presented as ICC (95% CI) for RD and KTW, Kappa (95% CI) for GT, CEJ, and RS, and quadratic weighted Kappa (95% CI) for RT.

Abbreviations: CEJ, cemento–enamel junction; CI, confidence interval; GT, gingival thickness; ICC, intraclass correlation coefficient; KTW, keratinized tissue width; RD, recession depth; RS, root step; RT, recession type.

counterparts (Tables 1 and 2). Inter-operator agreements were also very similar and within the same level of agreement for clinical and photographic measurements, except for RS and GT (Tables 1 and 2). RS showed substantial agreement for clinical measurements (Kappa = 0.72) but only moderate agreement for the photographic counterpart (Kappa = 0.45). It can be argued that the photographic representation of the lateral view of the tooth might not fully capture important details of the root conditions, reducing the accuracy of the detection of non-carious cervical lesions. Regarding GT, moderate inter-operator agreements were noted for clinical measurements (Kappa = 0.41), while only fair agreement was noted for photographic measurements (Kappa = 0.31). Of notice, GT is consistently associated with suboptimal agreement, both in clinical and photographic measurements, both for intra- and inter-operator agreements. However, the photographic inter-operator agreement for GT was lower than the clinical one. The flash could have contributed to the lower photographic agreement with altered color shades and confusing light marks. Also, an overview of the gingival phenotype of the patient might be unconsciously captured during the clinical examinations with views from different angles of the tooth, or from a view of different sites. This additional information was not available in the provided photographs. A more in-depth review of the indications and limitations of sulcus probing as an evaluation for GT was described by Pini Prato et al.[18] It should be noted that the problem of reduced agreement in photographic measurements, compared to clinical ones, does not seem to extend to other parameters, which showed similar levels of agreement between clinical and photographic evaluations. Interestingly, inter-operator agreement for KTW was slightly higher for photographic (ICC = 0.93) than for clinical evaluations (ICC = 0.88). This outcome might be explained by the presence of the probe aligned on the facial side captured in the picture, which allowed the standardization of photographic measurements. As for the clinical measurements, each operator positioned the probe independently during the diagnosis of RD and KTW, and yet, agreements remained very high.

Furthermore, measurements of the investigated parameters were found to influence the agreement for other parameters. Among the most interesting findings, RT showed a negative correlation with the agreement of RD in clinical measurements. This outcome implies that recession with a worse interproximal tissue level presented increased challenges for the accurate diagnosis of recession depth. The loss of papillae might have worsened the localization and measurement of midfacial recession, affecting the accuracy of RD measurements. As the interproximal tissue recedes, a wider area of the CEJ is exposed, creating challenges with a repeatable mesiodistal position of the probe and, consequently, different RD measurements. Interestingly, this negative correlation existed only for clinical measurements and not for photographic ones. Since photographic RD is measured on images containing a midfacial probe, examiners are not subject to variations induced by a different position of the probe during photographic evaluations. Another noteworthy correlation in measurement accuracy was found, indicating that KTW positively correlated with the accuracy of RD measurements, CEJ detectability, and the presence of RS, but only for clinical measurements. Clinical evaluation of oral tissues might provide a more comprehensive overview of the gingival phenotype and dental structure, creating interactions of variables in cases with more ambiguous presentations. As photographs provide only 1 viewpoint and timeframe of gingival recession, factors related to the gingival phenotype, including KTW, vestibular depth, and others, might have a lower likelihood of affecting photographic measurements. This is especially valid for RD, which is measured on a probe that exists within the picture.

Concerning photographic measurements, CEJ detectability was associated with the accuracy of RT and RS. It can be speculated that improved detectability of the CEJ helps clarify the position of the gingival margin

in relation to the CEJ, allowing for a more repeatable determination of RT. Noticeable CEJ also improves the localization of root structure, facilitating the identification of RSs or damages.

Moreover, agreement between clinical and photographic measurements resulted lowest for variables related to root conditions (CEJ and RS) (Table 3). One might speculate that frontal and lateral photographs don't adequately represent the tooth structure when compared to clinical examinations. During clinical examination, the operator is free to observe the tooth from different angles, to use explorer and periodontal probe to physically investigate the root surface. This infinite possibility of viewpoints and physical touch are absent in photographs, in which diagnosis is limited to frontal and lateral views, no physical exploration, reduced perception of depth of field, and unnatural lighting.

The present study has some limitations. Currently, there is a lack of information regarding the impact of various cameras and flash devices on the accuracy of photographic measurements. The effect of camera settings, such as color temperature and focus, on the reproducibility of photographic measurements remains unclear. It is important to note that, this study did not investigate whether the ethnic color of the gingiva might influence the reproducibility of GT, measured as the transparency of the probe during sulcus probing. This study maintains the limitations related to the use of a periodontal probe for investigating GRD, including rounding measurements to the nearest millimeter and variations in the position and angulation of the periodontal probe during clinical examinations. Several digital technologies have been identified to address the limitations of the periodontal probe in measuring gingival parameters.[23–26] The utilization of tridimensional scanning and digital rulers[23,24] or ultrasonography[27–30] allows for measurements with a precision near 0.001 mm. The decision to use a periodontal probe in this study was made to align with the 2018 Classification of GRD as outlined in the 2017 world workshop.[14,15] Additionally, the simpler the methodology of a study, the more easily it can be reproduced by other independent centers and utilized in clinical practice settings.

Telemedicine refers to the provision of healthcare services remotely using digital technology. It simplifies medical care access by allowing the sharing of digital records, reducing travel, barriers, and costs, and expanding provider reach. Telemedicine also presents advantages in research, particularly for multicenter studies involving centers located at a distance from each other.

Although telemedicine is widely utilized in medical and dental fields, its applications in periodontal plastic surgery are currently limited. To address this gap, a series of studies was designed to explore the integration of telemedicine concepts into the diagnosis of gingival recessions using the 2018 Classification of GRD. In the first multicenter study of this series, standardized photographic documentation of gingival recessions facilitated collaboration among 16 centers spanning 4 continents, conducting a measurement repeatability study based on the matrix of the 2018 Classification of GRD.[18] The variables associated with GRD demonstrated reproducibility within and between examiners. While the previous study exclusively used a digital diagnosis, using photographs as the only diagnostic tool, the current study compared 2 diagnostic methodologies, a clinical patient-based evaluation, and a digital method on photographs. In the current study, the repeatability of photographic measurements was compared to that of chairside measurements, revealing a similar level of agreement between the 2 methods. While this study validates the use of standardized photographs as an effective diagnostic tool in the field of periodontal plastic surgery, it is important to acknowledge that ongoing technological advancements provide a broad array of new technologies that allow the creation of high-resolution images, and increased precision of measurement.[23–28,30] Tools for image acquisition could be further enhanced by artificial intelligence-powered software for automated analysis and interpretation of the images. Such advancements would enable more accurate and consistent measurements, with important ramifications for therapy and follow-up monitoring.[31,32] These technologies need to be systematically tested against both clinical and photographic measurements to further enhance diagnostic capabilities in the field.

Future studies will investigate the application of photography for the diagnosis of gingival recession, and its possible use for telemedicine. Studies are encouraged from independent centers, using larger populations, with more heterogeneous ethnical distribution and testing different digital technologies.

5 | CONCLUSIONS

Within the limits of this study, it can be concluded that:

1. There is satisfactory intra-operator agreement for parameters constituting the 2018 Classification of gingival recession defect and gingival phenotype, both when analyzed chair-side and when evaluated in standardized photographs.
2. Satisfactory inter-operator agreement exists for RD, KTW, RS, and RT in clinical evaluations, while GT and CEJ exhibit lower inter-operator agreement. Additionally, satisfactory inter-operator agreement exists for RD, KTW, and RT in photographic evaluations, while

GT, CEJ, and RS demonstrate lower inter-operator agreement.
3. From substantial to excellent agreement was observed when comparing the clinical and photographic measurements, except for CEJ and RS.
4. Consistent evidence indicates that GT measured with sulcus probing has suboptimal agreement values, both in clinical and photographic measurements.

AUTHOR CONTRIBUTIONS

Riccardo Di Gianfilippo and GiovanPaolo Pini Prato conceived the study. Riccardo Di Gianfilippo, GiovanPaolo Pini Prato, Claudio M. Pannuti, and Leandro Chambrone designed the study. Francesco Cairo organized and controlled the clinical phases of the study. Francesco Cairo, Debora Franceschi, Luigi Barbato, and Walter Castelluzzo examined the cases and collected the data. Alessandra Bandel and Maria Di Martino contributed to the clinical phases of the study. Claudio M. Pannuti and Leandro Chambrone analyzed the data. Riccardo Di Gianfilippo, GiovanPaolo Pini Prato, Claudio M. Pannuti, Leandro Chambrone, and Francesco Cairo interpreted the data. Riccardo Di Gianfilippo and GiovanPaolo Pini Prato led the writing. All authors revised the manuscript before submission.

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. No external funding was available for this study.

FUNDING INFORMATION

No external funding was available for this study.

DATA AVAILABILITY STATEMENT

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

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

REFERENCES

1. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med*. 2008;121:S2-23.
2. Balogh EP, Miller BT, Ball JR. *Improving Diagnosis in Health Care*. National Academies Press; 2015:1-48.
3. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. *BMJ*. 2016;353:i2139. 2110.1136/bmj.i2139
4. Carbone AC, Joly JC, Botelho J, et al. Long-term stability of gingival margin and periodontal soft-tissue phenotype achieved after mucogingival therapy: a systematic review. *J Clin Periodontol*. 2023;51(2):177-195.
5. 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.
6. 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.
7. Nieri M, Rotundo R, Franceschi D, Cairo F, Cortellini P, Pini Prato G. Factors affecting the outcome of the coronally advanced flap procedure: a Bayesian network analysis. *J Periodontol*. 2009;80:405-410.
8. 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.
9. Cairo F, Cortellini P, Nieri M, et al. Coronally advanced flap and composite restoration of the enamel with or without connective tissue graft for the treatment of single maxillary gingival recession with non-carious cervical lesion. A randomized controlled clinical trial. *J Clin Periodontol*. 2020;47:362-371.
10. Sullivan HC, Atkins JH. Free autogenous gingival grafts. 3. Utilization of grafts in the treatment of gingival recession. *Periodontics*. 1968;6:152-160.
11. Miller PD Jr. A classification of marginal tissue recession. *Int J Periodontics Restorative Dent*. 1985;5:8-13.
12. 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.
13. 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.
14. Pini Prato G, Di Gianfilippo R. On the value of the 2017 classification of phenotype and gingival recessions. *J Periodontol*. 2021;92:613-618.
15. Cortellini P, Bissada NF. Mucogingival conditions in the natural dentition: narrative review, case definitions, and diagnostic considerations. *J Periodontol*. 2018;89(1):S204-S213.
16. Strauss FJ, Marruganti C, Romandini M, et al. Epidemiology of mid-buccal gingival recessions according to the 2018 classification system in South America: results from two population-based studies. *J Clin Periodontol*. 2023;50:1336-1347.
17. 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.
18. 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;94:661-672.
19. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527.
20. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med*. 1998;17:101-110.

21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
22. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155-163.
23. Rebele SF, Zuhr O, Schneider D, Jung RE, Hurzeler MB. Tunnel technique with connective tissue graft versus coronally advanced flap with enamel matrix derivative for root coverage: a RCT using 3D digital measuring methods. Part II. Volumetric studies on healing dynamics and gingival dimensions. *J Clin Periodontol*. 2014;41:593-603.
24. Zuhr O, Rebele SF, Schneider D, Jung RE, Hurzeler MB. Tunnel technique with connective tissue graft versus coronally advanced flap with enamel matrix derivative for root coverage: a RCT using 3D digital measuring methods. Part I. Clinical and patient-centred outcomes. *J Clin Periodontol*. 2014;41:582-592.
25. Windisch SI, Jung RE, Sailer I, Studer SP, Ender A, Hammerle CH. A new optical method to evaluate three-dimensional volume changes of alveolar contours: a methodological in vitro study. *Clin Oral Implants Res*. 2007;18:545-551.
26. Lehmann KM, Kasaj A, Ross A, Kammerer PW, Wagner W, Scheller H. A new method for volumetric evaluation of gingival recessions: a feasibility study. *J Periodontol*. 2012;83:50-54.
27. Barootchi S, Chan HL, Namazi SS, Wang HL, Kripfgans OD. Ultrasonographic characterization of lingual structures pertinent to oral, periodontal, and implant surgery. *Clin Oral Implants Res*. 2020;31:352-359.
28. Majzoub J, Tavelli L, Barootchi S, et al. Agreement in measurements of ultrasonography-derived periodontal diagnostic parameters among multiple raters: a diagnostic accuracy study. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2022;134:375-385.
29. Tavelli L, Majzoub J, Kauffmann F, et al. Coronally advanced flap versus tunnel technique for the treatment of peri-implant soft tissue dehiscences with the connective tissue graft: a randomized, controlled clinical trial. *J Clin Periodontol*. 2023;50:980-995.
30. Tavelli L, Yu N, Mancini L, Barootchi S. Keratinized mucosa width assessment at implant sites using high-frequency ultrasonography. *J Periodontol*. 2023;94:956-966.
31. Jain A, Way D, Gupta V, et al. Development and assessment of an artificial intelligence-based tool for skin condition diagnosis by primary care physicians and nurse practitioners in teledermatology practices. *JAMA Netw Open*. 2021;4:e217249.
32. Grossarth S, Mosley D, Madden C, et al. Recent advances in melanoma diagnosis and prognosis using machine learning methods. *Curr Oncol Rep*. 2023;25:635-645.

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: Di Gianfilippo R, Pini Prato G, Franceschi D, et al. Diagnostic reproducibility of the 2018 Classification of Gingival Recessions: Comparing photographic and in-person diagnoses. *J Periodontol*. 2025;96:467–477. https://doi.org/10.1002/JPER.24-0173