J Clin Periodontol 2014; 41: 927–933 doi: 10.1111/jcpe.12291

Journal of Clinical Periodontology

Inter-rater agreement in the diagnosis of mucositis and peri-implantitis

Mauro Merli[1,2,3], Francesco Bernardelli[1], Erica Giulianelli[1], Ivano Toselli[1], Marco Moscatelli[1], Umberto Pagliaro[4] and Michele Nieri[2,5]

[1]Private practice, Rimini, Italy; [2]INDENT, Rimini, Italy; [3]Politecnico delle Marche, Ancona, Italy; [4]Private practice, Florence, Italy; [5]Department of Surgery and Translational Medicine, University of Florence, Italy

Merli M, Bernardelli F, Giulianelli E, Toselli I, Moscatelli M, Pagliaro U, Nieri M. Inter-rater agreement in the diagnosis of mucositis and peri-implantitis. J Clin Periodontol 2014; 41: 927–933. doi: 10.1111/jcpe.12291.

Abstract

Aim: The objective was to assess the inter-rater agreement in the diagnosis of mucositis and peri-implantitis.

Material and Methods: Adult patients with ≥ 1 dental implant were eligible. Three operators examined the patients. One examiner allocated the patients to three groups of nine as follows: nine implants with peri-implantitis, nine implants with mucositis, and 9 implants with healthy mucosa. Each examiner recorded on all 27 patients (one implant per patient) recessions, probing depth, bleeding on probing, suppuration, keratinized tissue depth and bone loss, leading to a final diagnosis of mucositis, peri-implantitis or healthy mucosa. Examiners were independent and blinded to each other.

Results: Fleiss *k-statistic* with quadratic weight in the diagnosis of peri-implantitis and mucositis was 0.66 [CI95%: 0.45–0.87]. A complete agreement was obtained only in 14 cases (52%). Fleiss *k-statistics* in bleeding on probing and bone loss were respectively 0.31 [CI95%: 0.20–0.41] and 0.70 [CI95%: 0.45–0.94]. Intra-class correlation coefficients for recession, probing depth and keratinized tissue depth were respectively 0.69 [CI95%: 0.62–0.75], 0.54 [CI95%: 0.44–0.63] and 0.56 [CI95%: 0.27–0.77].

Conclusions: The inter-rater agreement in the diagnosis of peri-implant disease was qualified as merely good. This could also be due in part to the unclear definition of peri-implantitis and mucositis.

Key words: dental implant; diagnosis; inter-rater agreement; mucositis; peri-implantitis

Accepted for publication 14 July 2014

Implant loss is frequently due to peri-implantitis and appears to occur in multiple sites in afflicted patients (Roos-Jansåker 2007).

The Sixth European Workshop in Periodontics held in 2008 defined peri-implant diseases as follows: peri-implant mucositis is the presence of inflammation of the peri-implant mucosa without sign of loss of bone support, while peri-implantitis, in addition to inflammation of the mucosa, is characterized by a loss of bone support (Lindhe & Meyle 2008, Zitzmann & Berglundh 2008).

Peri-implantitis was often defined by the incidence of peri-implant probing depth ≥ 5 mm associated with bleeding on probing and/or suppuration and radiographic images of bone loss (Karoussis et al. 2004, Ferreira et al. 2006). Cases where the radiographs did not confirm the peri-implant bone loss were diagnosed as peri-implant mucositis (Costa et al. 2012).

Previous studies reported a high prevalence of peri-implant disease. Peri-implant mucositis occurs in about 80% of patients and 50% of implants (Lindhe & Meyle 2008). Peri-implantitis occurs in 28–56% of patients and 12–40% of implants (Lindhe & Meyle 2008).

Conflict of interest and source of funding statement

The authors declare that they have no conflict of interest. The study was funded by the authors.

In a Belgian sample study, the prevalence of mucositis and peri-implantitis at patient level were respectively 31% and 37% and were 38% and 23% at the implant level (Marrone et al. 2013). In another study, the prevalence of peri-implantitis was 7.5% at implant level (Maximo et al. 2008). The prevalence of mucositis and peri-implantitis varies in studies because different definitions of peri-implant disease are often used (Koldsland et al. 2010, Mombelli et al. 2012, Pesce et al. 2014).

The joint use of probing depth, radiographic bone loss and bleeding on probing was frequently implemented in the clinical diagnosis of peri-implant disease (Tomasi & Derks 2012). There are, however, noteworthy exceptions, for example in the 7th European Workshop on Periodontology peri-implantitis was characterized by changes in the level of the crestal bone in conjunction with bleeding on probing with or without concomitant deepening of peri-implant pockets (Lang & Berglundh 2011).

In a histologic study on dogs, the probing around implants represented a reliable technique for assessing the status of peri-implant mucosal health or disease (Lang et al. 1994).

In epidemiological studies, probing of periodontal pockets is the most commonly used method to quantify the periodontal status although numerous factors can affect inter-examiner variability and measurement accuracy (Holtfreter et al. 2012).

Two studies showed that peri-implant probing depth (PD) measurements are more sensitive to probing force variation than periodontal pocket probing (Mombelli et al. 1997, Gerber et al. 2009). The conclusion of a study on monkeys was that the probing measurements around osseo-integrated oral implants and teeth are different (Schou et al. 2002). Even mild marginal inflammation was associated with deeper probe penetration around implants in comparison to teeth (Schou et al. 2002). In conclusion, literature showed that PD measurement error was higher around implants than natural teeth (Eickholz et al. 2001).

A recent study on humans showed differences in pocket probing measurements at implants with or without the prosthetic reconstruction in place (Serino et al. 2013). The probing depth following prosthesis removal had a high correlation with the amount of bone loss at implants assessed during surgery (Serino et al. 2013).

A study showed that subjective evaluation of radiographs on simulated implants by skilled clinicians is rather uniform, and bone loss, follow-up and implant length are factors considered in the perception of implant success (Rotundo et al. 2011). Assessing radiographs of implant patients, the intra-observer agreement was good or very good (weighted kappa statistic between 0.72 and 0.82) while the inter-observer agreement was predominantly moderate (weighted kappa statistic between 0.58 and 0.62) (Kullman et al. 2007). The radiographic density and the distance between the reference point on the implant and the approximal marginal bone level had a marked and significant influence on the total inter-observer variation (Gröndahl et al. 1998). The darker the radiographs and the greater the distance, the larger the inter-observer variation (Gröndahl et al. 1998).

To date, no studies were conducted on the agreement between clinicians in the diagnosis of peri-implantitis in patients. The reliability of implant indexes as bleeding on probing and probing depth was rarely reported in literature (Verhoeven et al. 2000, Lachmann et al. 2007). In addition, the current definitions of peri-implantitis include pocket depth, bone loss, bleeding and suppuration and it is unclear how the disagreement in each of these factors contributes to the potential disagreement in the final diagnosis.

The objective of this study was to assess the inter-rater agreement in the diagnosis of mucositis and peri-implantitis. Agreement on probing depth, bleeding on probing, suppuration, keratinized tissue, bone loss and recession was also evaluated.

The manuscript was written following the guidelines for reporting reliability and agreement studies (GRRAS) (Kottner et al. 2011).

Material and Methods

Eligibility criteria for participants

The inclusion criteria were:

1 Eighteen years old or older.
2 Presence of ≥1 implant previously loaded for at least 1 year.
3 Current radiograph of the implant.

Gender, age and smoking status were recorded.

Informed consent

The investigators explained the nature of the trial, the aim and the methods to the patients, anticipating benefits, potential hazards as well as any form of discomfort that participation might entail. The patients read and asked questions inherent to the study prior to signing the informed consent. The informed consent was signed and dated by each patient before entering into the study. Patients were aware of their right to decline to participate or to withdraw from the study at any time.

Interventions

The procedure was performed in a private clinical centre specialized in implant therapy in Rimini (Italy) between November and December 2013. Three operators examined the patients. One operator had more than 12 years of experience in implant diagnosis and peri-implant therapy. The other two examiners had 4 and 5 years of experience in implant diagnosis and therapy. The examiner 1 chose nine implants in nine patients with peri-implantis, nine implants in nine patients with mucositis and nine implants in nine patients with healthy mucosa. The other two examiners measured the implants in random order. The interval between examinations was at least 20 min.

Primary outcome measurements

Prior to the start of the study examiner alignment was performed. Each examiner received a copy of the protocol and forms with the detailed measurement procedure. This allowed

the examiners to become familiar with the assessment process and clearly define the parameters indicated for peri-implantitis and mucositis. In an alignment meeting, the examiners' doubts were clarified and the procedure was explained.

In this study, peri-implantitis was defined as the presence of peri-implant probing depth ≥ 5 mm associated with bleeding on probing and/or suppuration at 1 peri-implant site with peri-implant radiographic bone loss (Karoussis et al. 2004). The definition of the 7th European Workshop on Periodontology was used as a sensitivity analysis: peri-implantitis is characterized by changes in the level of crestal bone in conjunction with bleeding on probing with or without concomitant deepening of peri-implant sites (Lang & Berglundh 2011).

Mucositis was defined as the presence of bleeding on probing without peri-implantitis.

Each examiner recorded recession, probing depth (PD), bleeding on probing (BoP), suppuration, keratinized tissue depth (KT) and bone loss, leading to a final diagnosis of mucositis, peri-implantitis or healthy mucosa on all 27 patients. Recession, probing depth, bleeding on probing and suppuration were measured at six sites per implant (mesio-vestibular, vestibular, disto-vestibular, disto-lingual, lingual, mesio-lingual). Gentle probing was applied using the PCPUNC15 probe (Hu-Friedy). An air syringe was used to dry the tissue prior to probing (Hefti & Preshaw 2012).

The keratinized tissue was registered at mid-buccal site.

Peri-implant bone loss was assessed in a dichotomous mode (yes or no). In the absence of previous radiographic records, the criterion used for the diagnosis of bone loss was based on a vertical distance threshold of 2 mm from the expected marginal bone level following remodelling post-implant placement (Sanz & Chapple 2012).

Blinding

The measurements were conducted independently and consecutively by the three examiners. The examiners were blinded to each other.

Sample size

The sample size calculation was performed using the function "CI3Cats" of the package "kappaSize" for R environment for statistical computing (Rotondi 2013). This function provides sample size estimation for *k statistic* using the confidence interval perspective and assumes that the outcome has three categories (Rotondi 2013). The anticipated value of *k* was set at 0.8, the lower bound of the CI95% was set at 0.6 and the upper bound at 0.99. The proportions of healthy implant, mucositis and peri-implantitis were set respectively at 0.33, 0.34 and 0.33 and the raters were set at 3. Using the previous parameters, a sample size of 25 subjects was necessary. To permit an equal number of the 3 conditions (healthy, mucositis, peri-implantitis) the final sample size was set to 27.

Statistical analysis

Intra-class correlation coefficients (ICC) and *k-statistics* with 95% confidence intervals were estimated to assess the inter-rater agreement of each variable. The *k*-statistic is a more robust measurement than a simple percented agreement calculation since it takes into account the agreement occurring by chance. If there is no agreement among the raters other than what is expected by chance, *k* is equal to 0.

Values equal or superior to 0.80 in *k* or ICC correspond to very good agreement; values between 0.61 and 0.80 signify good agreement; values between 0.41 and 0.60 correspond to moderate agreement; values between 0.21 and 0.40 signify fair agreement and values equal or inferior to 0.20 represent poor agreement (Altman 1991).

Fleiss *k* coefficient in the Gwet version using quadratic weight was used for diagnosis of peri-implant disease (Fleiss 1981, Gwet 2008).

As a sensitivity analysis for diagnosis of peri-implant disease, Gwet agree coefficient (AC_2) was used (Gwet 2012). Quadratic weights and custom weights were used. Based on the fact that clinical disagreement between healthy implants and mucositis is less important than the disagreement between healthy implants and peri-implantitis and between mucositis and peri-implantitis, custom weights were set at 0.25 to differentiate healthy implants and mucositis, and at 1 to differentiate healthy implants and peri-implantitis as well as mucositis and peri-implantitis. The same customized weights were used as a sensitivity analysis also using Fleiss *k* coefficient in the Gwet version.

As a secondary analysis, Bhapkar tests were performed pairwise between examiners to test for the homogeneity of the marginal distribution (Bhapkar 1966).

Intra-class correlation coefficients were calculated for recession and PD at site level and KT at implant level. The *k statistics* were calculated for bleeding on probing at site level and presence of peri-implant bone loss at implant level.

Two-way random intra-class correlation coefficients and *k-statistics* were reported paired between two examiners and combined among the three examiners. In the case of *k-statistic* between two examiners Cohen *k* with quadratic weights were used.

MedCalc® software version 12.7.5.0 (MedCalc Software, bvba, Ostend, Belgium) and AgreeStat software version 2013.1 (Advanced Analytic, LLC, Gaithersburg, MD) were used for the calculations.

Results

Twenty-seven patients were included in the study. One implant per patient was included and each patient completed the measurements with the three examiners.

The mean age of the patients was 59.2 years (ranged from 27 to 77 years), 15 (56%) were female, 7 (26%) were smokers and 17 implants (63%) were located in the maxillae.

Examiner 1 chose nine implants with peri-implantitis, nine implants with mucositis and nine healthy implants. In the expert diagnosis, only the nine implants with peri-implantitis had bone loss, KT was 2.0 mm (standard deviation 1.4 mm), recession was 0.2 mm (0.6 mm), PD was 3.7 mm (1.1 mm), bleeding on probing was present in 81 sites (50%) out of 162 and suppuration on probing was present in seven sites in six patients. Descriptive statistics for the

Table 1. Descriptive statistics

Variable	Examiner 1	Examiner 2	Examiner 3
Peri-implantitis	9 (33%)	6 (22%)	8 (30%)
Mucositis	9 (33%)	12 (44%)	13 (48%)
Healthy implant	9 (33%)	9 (33%)	6 (22%)
Bone loss	9 (33%)	6 (22%)	8 (30%)
Presence of suppuration	6 (22%)	0 (0%)	0 (0%)
Keratinized tissue (mm) (SD)	2.0 (1.4)	3.4 (1.6)	3.0 (1.7)
Bleeding on Probing [162 sites]	81 (50%)	86 (53%)	105 (65%)
Recession (mm) (SD) [162 sites]	0.2 (0.6)	0.2 (0.5)	0.2 (0.6)
PD (mm) (SD) [162 sites]	3.7 (1.1)	3.3 (1.2)	3.7 (1.2)

SD, Standard deviation; PD, Probing depth.

three examiners are reported in Table 1.

Fleiss *k-statistic* with quadratic weight in the diagnosis of peri-implantitis and mucositis was 0.66 [CI95% 0.45–0.87] (Table 2). Sensitivity analysis using Fleiss *k-statistics* with customized weights resulted in a value of 0.63 [CI95%: 0.41–0.85]. Gwet AC_2 with quadratic weight resulted in 0.71 [CI95%: 0.55–0.88] and Gwet AC_2 with customized weights resulted in 0.66 [CI95%: 0.46–0.86]. The values of Cohen *k-statistics* between two examiners are reported in Table 2.

According to the definition of the Consensus of the 7th European Workshop in Periodontology (Lang & Berglundh 2011) the results of this study are exactly the same.

A complete agreement was obtained only in 14 cases (52%): in five cases of peri-implantitis, five cases of mucositis and four cases of healthy implants. In the 13 disagreement cases, two examiners agreed and one examiner disagreed. There were no cases where all three examiners disagreed.

In eight cases, the difference in diagnosis was between mucositis and healthy implants (in 5 of these cases, one examiner diagnosed healthy implants despite the evidence of slight bleeding at one or two sites out of six implant sites). In four cases, the difference was between peri-implantitis and mucositis (no bone loss was recognized by one examiner in two cases and only one examiner recognized bone loss in the other two cases). In one case, two operators diagnosed peri-implantitis and one diagnosed healthy peri-implant tissue (no bone loss, one site with slight bleeding on probing and a PD = 4 mm were recognized by one examiner).

Bhapkar tests between raters were not significant ($p = 0.3713$ between examiner 1 and 2, $p = 0.2097$ between examiner 1 and 3 and $p = 0.2818$ between examiner 2 and 3).

Agreement on suppuration was not calculated since examiners 2 and 3 did not find suppurated sites.

Fleiss *k-statistic* for presence of bone loss at implant level was 0.70 [CI95% 0.45–0.94], and was 0.31 [CI95% 0.20–0.41] for presence of bleeding on probing at site level.

Intra-class correlation coefficient for recession and probing depth at site level were respectively 0.69 [CI95% 0.62–0.75] and 0.54 [CI95% 0.44–0.63]. Intra-class correlation coefficient for keratinized tissue depth at implant level was 0.56 [CI95% 0.27–0.77].

The *k-statistics* for presence of bone loss, presence of bleeding on probing and intra-class correlation coefficients for KT, recession and PD are shown paired between two examiners and combined among the three examiners in Table 3.

Table 2. Agreement on peri-implant disease: healthy implant, mucositis, peri-implantitis

Examiners	Kappa weighted	CI95%	Complete agreement
1 *versus* 2 *versus* 3**	0.66	0.45–0.87	14 (52%)
1 *versus* 2*	0.55	0.26–0.83	15 (56%)
1 *versus* 3*	0.81	0.66–0.96	21 (78%)
2 *versus* 3*	0.63	0.33–0.92	20 (74%)

*Cohen *k statistic* with quadratic weights; **Fleiss *k-statistic* with quadratic weights.

Discussion

The use of dental implants revolutionized the treatment of partially or totally edentulous patients in dentistry. This approach also determined the origin of new pathological conditions in peri-implant tissue, mucositis and peri-implantitis. Following the paradigms of periodontology, clinicians diagnosed peri-implant diseases based on peri-implant bone loss, bleeding on probing and probing depth.

The objective of this study was to assess inter-rater agreement in the diagnosis of mucositis and peri-implantitis. Agreement on probing depth, bleeding on probing, keratinized tissue, bone loss and recession were also evaluated.

Inconsistent data can yield serious consequences, such as false diagnoses and incorrect therapy, lack of discriminative power in clinical trials and excessive costs (Hefti & Preshaw 2012).

The result of the inter-rater agreement in this study for the diagnosis of peri-implant disease following the criteria proposed by Altman (1991) was merely good, the *k-statistic* was 0.66. Sensitivity analyses using different diagnostic criteria, different coefficient (Gwet AC_2) and customized weights did not substantially change the result. In addition, the statistical tests for homogeneity of the marginal distribution were not statistically significant.

The agreement on peri-implant probing depth (PD), recession, keratinized tissue depth (KT) and presence of bone loss were moderate or good, while the agreement on bleeding on probing (BoP) was fair. The agreement on suppuration was not calculated because examiner 2 and 3 did not find suppurated sites.

To date, to our knowledge, no previous study clinically calculated the inter-rater agreement in the diagnosis of peri-implant disease. In one study, inter-examiner weighted kappa coefficient for bone levels in intra-oral radiographs was 0.58 (Kullman et al. 2007). Similarly, in the present study *k-statistic* for presence of bone loss ranged from 0.55 to 0.91. In another study, inter-examiner ICC in PD in 85 peri-implant sites was 0.91 (Koldsland et al. 2010). In the

Table 3. Agreement on peri-implant factors

Variable \ Examiners	All [CI95%]	1 *versus* 2 [CI95%]	1 *versus* 3 [CI95%]	2 *versus* 3 [CI95%]
Presence of bone loss (k)	0.70 [0.45–0.94]	0.55 [0.19–0.90]	0.91 [0.80–1.00]	0.62 [0.27–0.97]
Keratinized tissue (ICC)	0.56 [0.27–0.77]	0.39 [−0.05–0.69]	0.57 [0.12–0.81]	0.76 [0.53–0.88]
Bleeding on Probing (162 sites) (k)	0.31 [0.20–0.41]	0.27 [0.12–0.42]	0.38 [0.25–0.52]	0.28 [0.14–0.43]
Recession (162 sites) (ICC)	0.69 [0.62–0.75]	0.68 [0.58–0.75]	0.70 [0.61–0.77]	0.68 [0.59–0.76]
PD (162 sites) (ICC)	0.54 [0.44–0.63]	0.49 [0.33–0.62]	0.56 [0.44–0.66]	0.57 [0.42–0.69]

k, *k-statistic*; ICC, Intra-class correlation coefficient; PD, Probing depth.

present study, the values for ICC related to peri-implant PD were lower, ranging from 0.49 to 0.57.

Previous *k-statistic* values for peri-implant BoP were not found in literature, although a study reported the inter-rater *k-statistics* for the gingiva index (Verhoeven et al. 2000). The inter-rater agreement on this variable was classified as fair in the present study, the *k-statistic* was only 0.31. Bleeding on probing is a potentially invasive process and a repeated assessment in a short period of time can lead to an overestimation of mucositis (Hefti & Preshaw 2012), which is yet to be proven. In this study, consecutive probing by the three operators resulted in a greater number of sites with bleeding during the last examination, hence, the diagnosis of mucositis following the previous two examinations could be attributed in part to this.

Peri-implant mucositis has been described as a disease in which the presence of inflammation is confined to the soft tissue surrounding a dental implant with no sign of loss of supporting bone following initial bone remodelling during healing (AAP 2013). From a clinical standpoint, signs of peri-implant mucositis include bleeding on probing and/or suppuration and no evidence of radiographic loss of bone beyond bone remodelling (AAP 2013).

Peri-implantitis has been characterized by an inflammatory process around an implant, which includes both soft tissue inflammation and progressive bone loss of supporting bone beyond biological bone remodelling (AAP 2013).

Although several thresholds were often indicated for probing depth and radiographic bone loss, and the criteria used to characterize peri-implant disease varied greatly between studies (Koldsland et al. 2010, Tomasi & Derks 2012, Pesce et al. 2014), the joint use of probing depth, radiographic bone loss and bleeding on probing was frequently implemented in the clinical diagnosis of peri-implant disease (Tomasi & Derks 2012).

In the present study, the examiners agreed with each other as to the definition of peri-implant disease prior to the start of the study. When examiners were unable to mutually define peri-implant disease, very poor agreement in diagnosis of this pathology resulted (Koldsland et al. 2010).

Definitions of peri-implant disease should be simple to use, accurate, reproducible, quantitative, quick and amenable to statistical analysis (Hefti & Preshaw 2012). In addition, a classification should be useful, exhaustive, disjointed and simple (Murphy 1997, Pini-Prato 2011).

The previous classifications of peri-implant disease are disjointed but not exhaustive. In fact, the combination of the three factors (PD, BoP and Bone loss) leads to 8 conditions. Only the condition with PD+ BoP+ BoneLoss+ leads to peri-implantitis, only the condition with PD− BoP− BoneLoss− leads to healthy implant and only the condition with PD− BoP+ BoneLoss− leads to mucositis.

Five combinations remain unclassified: PD- BoP+ BoneLoss+; PD+ BoP+ BoneLoss−, PD+ BoP− BoneLoss−, PD+ BoP− BoneLoss+, PD− BoP−BoneLoss+.

The combination PD− BoP+ BoneLoss+ could be frequent in cases of recessions. This condition cannot be considered peri-implantitis because the variable PD is negative and cannot be considered mucositis because there is a loss of supporting bone. Using the diagnostic criteria of the 7th European Workshop on Periodontology this combination can be classified as peri-implantitis.

The combination PD+ BoP+ BoneLoss- cannot be classified because of the possible presence of a pseudo-pocket or a pocket limited to the vestibular or lingual site of the peri-implant tissue.

The combinations of PD+ BoP- BoneLoss-, PD+ BoP− BoneLoss+, PD− BoP− BoneLoss+ could be present in cases of previous healed peri-implantitis.

The prevalence of the unclassified conditions is not known and could contribute to increasing disagreement. In the present study, the combinations of the unclassified PD− BoP+ BoneLoss+ and PD+ BoP+ BoneLoss− were not observed by the three examiners and for this reason the two classifications used yield exactly the same results.

Furthermore, incongruity may arise from the actual diagnostic process, wherein peri-implant disease is defined at implant level but diagnosis is derived at site level. For example, an implant that shows PD+ BoP− BoneLoss+ at a mesial site (unclassified, perhaps a previously healed site) and at the same time registers PD− BoP+ BoneLoss− at a distal site (classified as mucositis) could technically indicate peri-implantitis (PD+ BoP+ BoneLoss+) at implant level. Another situation can arise when only one site is slightly bleeding. In the present study, this result was frequently considered healthy by the raters in an attempt to reduce the false positive rate.

BoP has considerable value in the classification of periodontal disease. It is unclear if this factor can predict future bone loss, but in the present study the agreement on this variable is fair. The BoP measurement can change the diagnosis from a healthy implant to mucositis, as well as from an unclassified situa-

tion (PD+ BoP− BoneLoss+) to peri-implantitis.

New diagnostic tools, such as biochemical indicators of peri-implant sulcular fluid (the matrix metalloproteinase 8) could facilitate the diagnosis and the inter-rater agreement of peri-implant diseases (Kivelä-Rajamäki et al. 2003, Basegmez et al. 2012).

The intra-rater agreement was not performed in this study. The intention was to inspect only the inter-rater agreement in the diagnostic procedure.

A limit of the present study is that it was conducted in a private setting by dental professionals working there on a daily basis which could have introduced a bias due to the examiners familiarity with the study subjects. However, it is important to note that the study centre treats several thousand patients per year, the diagnoses were not performed prior to the study initiation and were performed independently. In addition, the examiners were blinded to each other.

Another limit of the present study is that the k value assumed in the sample size calculation was different from the k estimated in the results. Nevertheless, the precision of the estimates, namely the width of the confidence intervals of the results are very similar to the confidence interval width preset in the sample size calculation.

In summary, the inter-rater agreement in the diagnosis of peri-implant disease was qualified as merely good. This could also be due in part to the unclear definition of peri-implantitis and mucositis.

Acknowledgements

The authors express their deepest appreciation to Heather Dawe for editing the manuscript.

References

American Academy of Periodontology (AAP) (2013) Peri-implant mucositis and peri-implantitis: a current understanding of their diagnoses and clinical implications. *Journal of Periodontology* **84**, 436–443.

Altman, D. G. (1991) *Practical statistics for medical research*, p. 404. London, England: Chapman & Hall/CRC.

Basegmez, C., Yalcin, S., Yalcin, F., Ersanli, S. & Mijiritsky, E. (2012) Evaluation of periimplant crevicular fluid prostaglandin E2 and matrix metalloproteinase-8 levels from health to peri-implant disease status: a prospective study. *Implant Dentistry* **21**, 306–310.

Bhapkar, V. P. (1966) A note on the equivalence of two test criteria for hypotheses in categorical data. *Journal of the American Statistical Association* **61**, 228–235.

Costa, F. O., Takenaka-Martinez, S., Cota, L. O., Ferreira, S. D., Silva, G. L. & Costa, J. E. (2012) Peri-implant desease in subject with and without preventive maintenance: a 5-year follow-up. *Journal of Clinical Periodontology* **39**, 173–181.

Eickholz, P., Grotkamp, F. L., Steveling, H., Mühling, J. & Staehle, H. J. (2001) Reproducibility of peri-implant probing using a force-controlled probe. *Clinical Oral Implants Research* **12**, 153–158.

Ferreira, S. D., Silva, G. L., Cortelli, J. R., Costa, J. E. & Costa, F. O. (2006) Prevalence and risk variables for peri-implant disease in Brazilian subjects. *Journal of Clinical Periodontology* **33**, 929–935.

Fleiss, J. L. (1981) *Statistical methods for rates and proportions*, 2nd edition, pp. 211–236. New York: John Wiley & Sons.

Gerber, J. A., Tan, W. C., Balmer, T. E., Salvi, G. E. & Lang, N. P. (2009) Bleeding on probing and pocket probing depth in relation to probing pressure and mucosal health around oral implants. *Clinical Oral Implants Research* **20**, 75–78.

Gröndahl, K., Sundén, S. & Gröndahl, H. G. (1998) Inter- and intraobserver variability in radiographic bone level assessment at Brånemark fixtures. *Clinical Oral Implants Research* **9**, 243–250.

Gwet, K. L. (2008) Computing inter-rater reliability and its variance in the presence of high agreement. *The British Journal of Mathematical and Statistical Psychology* **61** , 29–48.

Gwet, K. L. (2012) *Handbook of inter-rater reliability*, 3rd edition, pp. 89–92. Gaithersburg: Advanced Analytics, LLC.

Hefti, A. F. & Preshaw, P. M. (2012) Examiner alignment and assessment in clinical periodontal research. *Periodontology 2000* **59**, 41–60.

Holtfreter, B., Alte, D., Schwahn, C., Desvarieux, M. & Kocher, T. (2012) Effects of different manual periodontal probes on periodontal measurements. *Journal of Clinical Periodontology* **39**, 1032–1041.

Karoussis, I. K., Brägger, U., Salvi, G. E., Bürgin, W. & Lang, N. P. (2004) Effect of implant design on survival and success rates of titanium oral implants: a 10-year prospective cohort study of the ITI Dental Implant System. *Clinical Oral Implants Research* **15**, 8–17.

Kivelä-Rajamäki, M., Maisi, P., Srinivas, R., Tervahartiala, T., Teronen, O., Husa, V., Salo, T. & Sorsa, T. (2003) Levels and molecular forms of MMP-7 (matrilysin-1) and MMP-8 (collagenase-2) in diseased human peri-implant sulcular fluid. *Journal of Periodontal Research* **38**, 583–590.

Koldsland, O. C., Scheie, A. A. & Aass, A. M. (2010) Prevalence of peri-implantitis related to severity of the disease with different degrees of bone loss. *Journal of Periodontology* **81**, 231–238.

Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B. J., Hróbjartsson, A., Roberts, C., Shoukri, M. & Streiner, D. L. (2011) Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Journal of Clinical Epidemiology* **64**, 96–106.

Kullman, L., Al-Asfour, A., Zetterqvist, L. & Andersson, L. (2007) Comparison of radiographic bone height assessments in panoramic and intraoral radiographs of implant patients. *The International Journal of Oral & Maxillofacial Implants* **22**, 96–100.

Lachmann, S., Kimmerle-Müller, E., Axmann, D., Gomez-Roman, G., Weber, H. & Haas, R. (2007) Reliability of findings around healthy implants in association with oral hygiene measures: a clinical, microbiological, and immunological follow-up in edentulous patients. *Clinical Oral Implants Research* **18**, 686–698.

Lang, N. P., Berglundh, T. & on behalf of Working Group 4 of the Seventh European Workshop on Periodontology (2011) Periimplant diseases: where are we now?–Consensus of the Seventh European Workshop on Periodontology. *Journal of Clinical Periodontology* **38** (Suppl. 11), 178–181.

Lang, N. P., Wetzel, A. C., Stich, H. & Caffesse, R. G. (1994) Histologic probe penetration in healthy and inflamed peri-implant tissues. *Clinical Oral Implants Research* **5**, 191–201.

Lindhe, J., Meyle, J., & Group D of European Workshop of Periodontology (2008) Peri-implant deseases: Consensus Report of the Sixth European Workshop on Periodontology. *Journal of Clinical Periodontology* **35** (8 Suppl.), 282–285.

Marrone, A., Lasserre, J., Bercy, P. & Brecx, M. C. (2013) Prevalence and risk factor for peri-implant disease in Belgian adults. *Clinical Oral Implants Research* **24**, 934–940.

Maximo, M. B., de Mendonca, A. C., Alves, J. F., Cortelli, S. G., Peruzzo, D. G. & Duarte, P. M. (2008) Peri-implant diseases may be associated with increased time loading and generalized periodontal bone loss: preliminary results. *The Journal of Oral Implantology* **34**, 268–273.

Mombelli, A., Mühle, T., Brägger, U., Lang, N. P. & Bürgin, W. B. (1997) Comparison of periodontal and peri-implant probing by depth-force pattern analysis. *Clinical Oral Implants Research* **8**, 448–454.

Mombelli, A., Müller, N. & Cionca, N. (2012) The epidemiology of peri-implantitis. *Clinical Oral Implants Research* **23** (Suppl. 6), 67–76.

Murphy, E. A. (1997) *The logic of medicine*, 2nd edition. Baltimore and London: The John Hopkins University Press.

Pesce, P., Menini, M., Tealdo, T., Bevilacqua, M., Pera, F. & Pera, P. (2014) Peri-implantitis: a systematic review of recently published papers. *International Journal of Prosthodontics* **27**, 15–25.

Pini-Prato, G. (2011) The Miller classification of gingival recession: limits and drawbacks. *Journal of Clinical Periodontology* **38**, 243–245.

Roos-Jansåker, A. M. (2007) Long time follow-up of implant therapy and treatment of peri-implantitis. *Swedish Dental Journal* **188** (Suppl.), 7–66.

Rotondi, M. A. (2013) Package 'kappaSize'. URL http://cran.r-project.org/web/packages/kappaSize/kappaSize.pdf.

Rotundo, R., Nieri, M., Briccoli, L., Barone, R. & Clauser, C. (2011) Radiographic interpretation agreement in implant dentistry. *European Journal of Oral Implantology* **4**, 135–143.

Sanz, M., Chapple, I. L. & on behalf of Working Group 4 of the VIII European Workshop on Periodontology (2012) Clinical research on peri-implant diseases: consensus report of Working Group 4. *Journal of Clinical Periodontology* **39** (Suppl. 12), 202–206.

Schou, S., Holmstrup, P., Stoltze, K., Hjørting-Hansen, E., Fiehn, N. E. & Skovgaard, L. T. (2002) Probing around implants and teeth with healthy or inflamed peri-implant mucosa/gingiva. A histologic comparison in cynomolgus monkeys (Macaca fascicularis). *Clinical Oral Implants Research* **13**, 113–126.

Serino, G., Turri, A. & Lang, N. P. (2013) Probing at implants with peri-implantitis and its relation to clinical peri-implant bone loss. *Clinical Oral Implants Research* **24**, 91–95.

Tomasi, C. & Derks, J. (2012) Clinical research of peri-implant diseases–quality of reporting, case definitions and methods to study incidence, prevalence and risk factors of peri-implant diseases. *Journal of Clinical Periodontology* **39** (Suppl. 12), 207–223.

Verhoeven, J. W., Cune, M. S. & de Putter, C. (2000) Reliability of some clinical parameters of evaluation in implant dentistry. *Journal of Oral Rehabilitation* **27**, 211–216.

Zitzmann, N. U. & Berglundh, T. (2008) Definition and prevalence of peri-implant disease. *Journal of Clinical Periodontology* **35** (8 Suppl.), 286–291.

Address:
Mauro Merli
Private practice
Viale Settembrini 17/O 47923 Rimini
Italy
E-mail: mauromerli@gmail.com

Clinical Relevance

Scientific rationale for the study: There has been little published regarding the inter-rater agreement between dental operators in the diagnosis of peri-implant disease and in measuring peri-implant variables.

Principal findings: The inter-rater agreement in diagnosis of peri-implant disease was merely good (*k-statistic* was 0.66). The agreement on peri-implant probing depth, recession, keratinized tissue and presence of bone loss was moderate or good while the agreement on bleeding on probing was fair.

Practical implications: Establishing a consensus on this subject could provide guidelines for early diagnosis of borderline cases. A better agreement could result in programs to improve the implant supported therapy.