Journal of Dental Sciences (2021) 16, 948–956

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.e-jds.com

Journal of Dental Sciences

Original Article

Intra- and inter-observer agreements in detecting peri-implant bone defects between periapical radiography and cone beam computed tomography: A clinical study

Chu-Nan Zhang [a], Yu Zhu [a], Lin-Feng Fan [b], Xiao Zhang [a], Yin-Hua Jiang [c], Ying-Xin Gu [a*]

[a] *Department of Implant Dentistry, Shanghai Ninth People's Hospital, College of Stomatology, Shanghai Jiao Tong University School of Medicine; National Clinical Research Center for Oral Diseases; Shanghai Key Laboratory of Stomatology & Shanghai Research Institute of Stomatology, Shanghai, China*

[b] *Department of Radiology, Shanghai Ninth People's Hospital, College of Stomatology, Shanghai Jiao Tong University School of Medicine, Shanghai, China*

[c] *Department of Oral and Maxillofacial Implantology, Lishui Sixth People's Hospital, Wenzhou Medical University, Lishui, China*

Received 15 October 2020; Final revision received 28 October 2020
Available online 14 November 2020

KEYWORDS
Cone beam computed tomography;
Dental implants;
Periapical radiography;
Peri-implant bone defect

Abstract *Background/purpose:* Information regarding agreements between periapical radiograph (PA) and cone beam computed tomography (CBCT) in detecting peri-implant defect is still scarce. The aim of this clinical study was to compare agreements between PA and CBCT in detecting peri-implant bone defect.

Materials and methods: This retrospective clinical study enrolled 32 patients with both PA and CBCT filmed right after implant placement. Four modalities were used for film reading: PA1 (original), PA2 (enhanced brightness/contrast), CBCT1 (selected axial and mesial-distal direction images) and CBCT2 (all data with software). 2 experienced and 2 inexperienced observers scored all films. Intra- and inter-observer agreements were estimated with Cohen's kappa coefficient. Categorized agreements were compared and differences among four modalities were calculated.

Results: Agreements of PA were better than CBCT when detecting peri-implant bone defects in inter-observer agreements (median kappa 0.471 vs. 0.192; $p = 0.016$). Moreover, agreements in experienced observers were better than inexperienced observers (median kappa 0.883 vs. 0.567; $p < 0.001$). There was significant difference among four modalities except for experienced observer 2 ($p = 0.218$).

* Corresponding author. Department of Implant Dentistry, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, 639 Zhizaoju Road, Shanghai, 200011, China. Fax: +86 21 53073068.
E-mail address: yingxingu@163.com (Y.-X. Gu).

https://doi.org/10.1016/j.jds.2020.10.013

Conclusion: Agreements of PA are better than CBCT when detecting peri-implant bone defects, especially for inter-observer agreements. Experienced observers are more consistent in assessment than inexperienced ones.

Introduction

Radiological examination is crucial in pre- and post-operative assessment of implant surgery. Peri-implant bone defects may be found in implants with surgical trauma, peri-implantitis and osseointegration failure. Among all radiological methods, periapical radiography (PA) and cone beam computed tomography (CBCT) are common ways to assess post-operative peri-implant bone defects.

Periapical radiography is considered as an accurate and reliable tool to assess peri-implant bone status with favorable costs and radiation dose. However, the image of PA is only two-dimensional with superimposed structures. Since CBCT has been introduced to dentistry, it enables clinicians to obtain 3-dimensional images with a low dose of radiation. CBCT has been widely used in implant dentistry, including pre-implant assessment of anatomy and post-implant placement.[1] One potential limitation of CBCT in evaluating post-operative peri-implant status is the presence of metal artifacts, which could mask osseointegration, shallow bony defects and other peri-implant radiolucencies,[2] thus may jeopardize diagnosis.

There are many researches focusing on comparison between PA and CBCT at detecting peri-implant bone defects. One review article stated that there was no difference between PA and CBCT in detecting peri-implant bone.[3] However, agreements between PA and CBCT in detecting peri-implant bone defect are inconsistent. Studies conducted by Dave et al. and Vidor et al. proved that PA was a reliable tool of detecting peri-implant bone defects and performed significantly better than CBCT in agreements.[4,5] However, other studies showed that agreements of CBCT were better than of PA when assessing bone defect around implants.[6–8] To date, information regarding agreements between PA and CBCT in detecting peri-implant defect is still scarce.

Moreover, most of the previous studies were designed in vitro. Little is known about defect detecting agreement between PA and CBCT in clinical situation. Furthermore, some issues still require an appropriate discussion: different PA film brightness/contrast, different CBCT reading methods and influence of expertise of observer.

Therefore, the aim of the present study was to compare agreements between PA and CBCT in bone defect assessment around dental implants in clinical situation. Our hypothesis was that agreements for PA and CBCT would be comparable.

Materials and methods

The present study was approved by the Research Ethics Committee for clinical study, Shanghai Jiao Tong University, under protocol No. 2016-236-C33. A total 44 patients were recruited and all participants have signed informed consent. This clinical study adheres to the principles described by the Declaration of Helsinki that revised in 2013 for research involving human participants. All methods performed were in accordance with the relevant guidelines and regulations. After consultation, individuals who met the following inclusion criteria were asked to participate in the study:

(a) ≥18 years of age
(b) no systematic or local conditions presenting a contraindication to implant placement
(c) requirement of single implant placement with guided bone regeneration
(d) no pathologic conditions of teeth adjacent to the area of surgical procedure
(e) periodontium in the area of the surgical procedure to be free of pathology and artifacts such as pins and wires

The following exclusion criteria were also applied:

(a) presence of complex systematic disease or bone metabolic disorder
(b) history of malignancy, radiotherapy, or chemotherapy in the past 5 years
(c) smoking > 10 cigarettes
(d) pregnancy or lactating woman
(e) no attempt to sign consent form

Surgical procedure

A total of 34 patients were enrolled in this study and all informed consents were obtained prior to surgery. All patients received single implant placement with guided bone regeneration. A two-piece, tapered implant system was used (ETGAR medical implant systems, Nahariya, Israel). Bovine bone matrix composed of hydroxyapatite collagen and collagen membrane (Zhenghai Biotechnology, Shandong, China) were used for guided bone regeneration. CBCT and PA were both filmed right after implant placement.

Image acquisition

The periapical films were taken using a size 31 mm × 41 mm digital photostimulable phosphor plate (PSP) (Digora Optime, Soredex, Tuusula, Finland) with a standard intraoral X-ray machine (Heliodent DS, Sirona, Bensheim, Germany) with exposure settings of 70 kV, 7 mA for 0.08 s.

CBCT was undertaken using i-CAT 17–19 (Imaging Sciences International, Hatfield, PA, USA). The exposure settings selected were: 360°scan, 120 kV, 5 mA and acquisition time 14.7 s. The scanning parameters were FOV height*-width 8 cm × 8 cm and voxel size 0.25 mm.

Image preparation

In total, all PA and CBCT of 34 patients were acquired. After carefully screening, 2 CBCT images with motion artifacts were excluded. Finally, 32 patients were enrolled in this study. No prosthetic crowns or root canal filling materials were seen around implants.

Two researchers, who didn't involve in image analysis, were in charge of preparing all films. For PA, two files were created. The first file, named PA1, contained the original digital films. The second file, named PA2, was films adjusted with increasing brightness/contrast subjectively by agreement between those two researchers. For CBCT, two reading methods were used. The first method was to create one file, named CBCT1, with two images selected from software per patient. An axial image (around apical third of implants) and a mesial-distal direction image (taken nearly through the implant center in the axial views) were selected subjectively by agreement between those two researchers. Thus, there were two images per patient, which were considered to be the most representative of this patient, for observers to evaluate peri-implant bone status. These images were chosen so all observers could carry out evaluation on identical images. The second method CBCT2 was to scroll through all data by using software, and to change the contrast and brightness of each patient when needed. Observers were allowed to perform any section in multi-planar reconstructions (MPR) to evaluate peri-implant bone status. The order of the images in the above four files was scrambled and anonymization of all images was performed. Instructions for reading PA and CBCT were also prepared by the two researchers.

Image evaluation

PowerPoint instructions, files of PA films and CBCT were prepared and stored in identical device. Observers were asked to review the radiographs on the same electronic device under quiet, darkened environmental condition. The sequence of periapical films and CBCT images was randomized using Microsoft Excel (Microsoft Excel for Mac 2011, Microscope Corp, Seattle, WA, USA). The viewing and adjusting software for images was Microsoft PowerPoint (Microsoft PowerPoint for Mac 2011, Microscope Corp) and for CBCT discs was i-CAT Vision (Imaging Sciences International).

Four observers consisting of two dentists with more than 5 years' experience in implant imaging (Observer 1 and Observer 2) and two postgraduate students majored in implant dentistry (Observer 3 and Observer 4) were asked to evaluate all radiographs. All observers were required to repeat the entire evaluation 2 weeks later to produce data for intra-observer reliability. The image order changed for the second evaluation.

A 5-point scale was used, with the following classification:

1. Peri-implant bone defect definitely not present
2. Peri-implant bone defect probably not present
3. Unsure if peri-implant bone defect present or not
4. Peri-implant bone defect probably present
5. Peri-implant bone defect definitely present.

For every patient, both mesial and distal bone status of implant were evaluated. All data were transferred to Excel sheets (Microsoft Excel for Mac 2011, Microscope Corp) for analysis.

Statistical analysis

The data were analyzed using Statistical Analysis System, SAS version 9.4 (SAS Institute, Cary, NC, USA). All acquired scores from all observers were summarized. Then, score 1,2,3 were pooled as "0", indicating absence of bone defect, while score 4,5 were pooled as result "1", indicating presence of bone defect. Pooled results for each observer were analyzed for intra-observer and inter-observer agreement. Agreements were analyzed by the pooled results of mesial and distal score, respectively. For example, if mesial and distal scores of a patient were 4 and 3. Then agreement was analyzed by the pooled result of "1" and "0". Cohen's Kappa analysis used for all comparisons between observers was seen in Table 1.

After all kappa results acquired, sorted value were categorized. Wilcoxon signed rank test (for paired data) was used to compare differences between intra-observer PA and CBCT, inter-observer PA and CBCT, experienced and inexperienced observers, PA1 and PA2, as well as CBCT1 and CBCT2. In this study, any kappa values under 0.4 were regarded as poor agreement, between 0.40 and 0.59 were regarded as moderate agreement, any over 0.6 were considered as good agreement and any over 0.8 as excellent agreement (Landis & Koch 1977).

Averaged results of two assessments in each patient were used for comparison among different imaging modalities. Cochran's Q-test with a post hoc McNemar test and Benjamini and Hochberg correction were used to test for statistically significant differences between the four imaging modalities. Then, all averaged results from all 4 observers with all 4 imaging modalities for each patient added up. For example, if all results were "1", the highest score one could get was "16". Then, total score of each patient was ranked with proportion of scores from CBCT calculated. Spearman correlation between final score and proportion of CBCT was analyzed afterwards.

P value smaller than 0.05 was regarded as statistically significant. As the McNemar test involved four comparisons, a correction for multiple tests was carried out and statistical significance was accepted at $p < 0.01$.

Results

All four observers completed all radiographic assessments. Ages of 32 patients ranged from 21 to 60, with 14 females and 18 males. 1 anterior tooth and 31 posterior teeth were replaced by implants. Four film-reading modalities were used with twice assessment both at mesial and distal aspects. Thus, each observer had 512 radiographic assessments.

Initial results were outlined in Table 2. Observer 1 and 2 were experienced observers, while observer 3 and 4 were graduate students. The results showed variation among different observers. Obviously, the number of score "3" in experienced observers was much less than in graduate

Table 1 Cohen's Kappa variables used for comparisons between observers.

Cohen's kappa for intra- and inter-observer agreements were calculated for the following comparisons		
Intra-observer agreement for all observers	PA1-1 versus PA1-2	The first and second assessments with original periapical film
	PA2-1 versus PA2-2	The first and second assessments with increased brightness/contrast of periapical film
	CBCT 1-1 versus CBCT 1-2	The first and second assessments with CBCT selected images
	CBCT 2—1 versus CBCT 2-2	The first and second assessments with CBCT all data
Inter-observer agreement for averaged PA1, PA2, CBCT1, CBCT2	O1 versus O2 (PA1) O3 versus O4 (PA1)	Experienced and inexperienced observer results for original periapical film
	O1 versus O2 (PA2) O3 versus O4 (PA2)	Experienced and inexperienced observer results for periapical film with increased brightness/contrast
	O1 versus O2 (CBCT1) O3 versus O4 (CBCT1)	Experienced and inexperienced observer results for CBCT with selected images
	O1 versus O2 (CBCT2) O3 versus O4 (CBCT2)	Experienced and inexperienced observer results for CBCT with all data

O, Observer; PA1, original periapical radiography; PA2, enhanced brightness/contrast periapical radiography; CBCT1, selected axial and mesial-distal direction images; CBCT2, all data with software.

students. After pooling, numbers of "1" were least in observer 1 and most in observer 3. Moreover, in experienced observers, the number of "1" was more in PA than in CBCT, while in graduate students, the number of result "1" was more in CBCT.

The kappa values for intra- and inter-observer agreements were summarized in Table 3. For experienced observers, intra-observer agreements of PA and CBCT at both levels were mostly "excellent" (range 0.795—1.000). In general, kappa values for PA2 and CBCT2 were slightly better than PA1 and CBCT1, respectively. However, agreements for observer 3 and 4 were mostly "moderate to good" (range 0.410—1.000). Meanwhile, kappa values for PA2 were not always better than PA1, while agreements of CBCT2 were better than CBCT1.

For experienced observers, inter-observer agreements at both levels were mostly "poor to moderate", with poor for CBCT and moderate for PA. Kappa values for PA2 and CBCT2 were slightly better than PA1 and CBCT1, respectively. However, inter-observer agreements for observer 3 and 4 were not as uniform as for observer 1 and 2. One inter-observer values for observer 3 and 4 were noticed, for result was below poor (mesial CBCT2 -0.0303). All intra- and inter-agreements were summarized in Fig. 1.

Results from Wilcoxon signed rank test was shown in Table 4. Significant differences were seen in inter-observer PA/CBCT and experienced/inexperienced observers. Results of Cochran's test were summarized in Table 5. It showed that there was significant difference among four modalities except for observer 2. Significant difference was seen in observer 1, observer 3 and observer 4. However, greater Q values were only seen in observer 3 and observer 4. Moreover, post-hoc multiple tests showed that PA2CBCT1 and PA2CBCT2 of observer 3 and PA2CBCT2 of observer 4 were the most prominent difference.

The Spearman's correlation was −0.74 (Fig. 2). With total score of each patient decreasing, proportion of scores from CBCT was increasing. Moreover, there were 3 patients whose results were all "1" with total score "16". The lowest final score was "2" in two patients, which were all contributed from CBCT. Images of these patients were seen in Fig. 3. Three patients with peri-implant bone defect identified by all 4 observers were on the left, while two patients with lowest scores were on the right.

Discussion

Agreements in detecting peri-implant bone defects of PA were better than CBCT, especially for inter-observer agreements. Thus, our postulated hypothesis, that was, agreements for PA and CBCT would be comparable, was

Table 2 Summary of initial score answers for each observer.

Observer	Methods	Scores				
		1	2	3	4	5
1	PA1	0	63	30	14	21
	PA2	0	66	26	19	17
	CBCT1	0	74	31	7	16
	CBCT2	0	80	25	3	20
2	PA1	2	38	29	31	28
	PA2	2	40	18	39	29
	CBCT1	16	45	13	41	24
	CBCT2	6	40	23	35	25
3	PA1	25	35	0	35	33
	PA2	37	38	0	29	24
	CBCT1	15	30	7	42	34
	CBCT2	3	29	11	42	43
4	PA1	31	46	3	28	20
	PA2	46	48	3	18	13
	CBCT1	33	47	5	22	21
	CBCT2	30	30	3	23	42

PA1, original periapical radiography; PA2, enhanced brightness/contrast periapical radiography; CBCT1, selected axial and mesial-distal direction images; CBCT2, all data with software.

Table 3 Intra- and inter-observer agreements for each observer (95% confidence interval).

Observer	Methods	Level1-mesial	Level1-distal
1	PA1	0.932 (0.802–1.000)	0.795 (0.522–1.000)
	PA2	0.932 (0.802–1.000)	0.904 (0.719–1.000)
	CBCT1	0.796 (0.528–1.000)	0.890 (0.680–1.000)
	CBCT2	0.904 (0.719–1.000)	1.000 (1.000–1.000)
2	PA1	0.937 (0.816–1.000)	0.867 (0.691–1.000)
	PA2	1.000 (1.000–1.000)	1.000 (1.000–1.000)
	CBCT1	0.875 (0.709–1.000)	0.938 (0.817–1.000)
	CBCT2	0.937 (0.816–1.000)	0.934 (0.808–1.000)
3	PA1	0.629 (0.370–0.889)	0.629 (0.370–0.889)
	PA2	0.431 (0.190–0.671)	0.529 (0.259–0.801)
	CBCT1	0.410 (0.087–0.732)	0.552 (0.260–0.844)
	CBCT2	0.503 (0.161–0.844)	0.688 (0.449–0.927)
4	PA1	0.563 (0.290–0.835)	0.649 (0.374–0.924)
	PA2	0.600 (0.286–0.915)	1.000 (1.000–1.000)
	CBCT1	0.570 (0.242–0.898)	0.625 (0.363–0.887)
	CBCT2	0.688 (0.436–0.939)	0.751 (0.524–0.978)
O1 versus O2	PA1	0.459 (0.186–0.732)	0.363 (0.072–0.664)
	PA2	0.582 (0.334–0.831)	0.482 (0.216–0.748)
	CBCT1	0.073 (−0.171–0.316)	0.125 (0.143–0.393)
	CBCT2	0.358 (0.122–0.593)	0.321 (0.011–0.630)
O3 versus O4	PA1	0.625 (0.374–0.876)	0.290 (−0.010–0.591)
	PA2	0.458 (0.184–0.731)	0.739 (0.462–1.000)
	CBCT1	0.189 (0.034–0.412)	0.343 (0.055–0.632)
	CBCT2	−0.030 (−0.321–0.261)	0.194 (0.139–0.527)

O, Observer; PA1, original periapical radiography; PA2, enhanced brightness/contrast periapical radiography; CBCT1, selected axial and mesial-distal direction images; CBCT2, all data with software.

rejected. As we know, this is the first clinical study assessing agreements between PA and CBCT and PA showed a more favorable diagnostic consistency than CBCT in detecting peri-implant bone defects. Clinical variation in diagnosing defects around dental implants should be noted in clinical practice.

In fact, there were only a few in vitro studies comparing agreements between PA and CBCT and results were inconsistent. For example, Schwindling et al. and Hilgenfeld et al. created at least 1 mm diameter 1 wall to 4 walls peri-implant bone defects and results showed that agreements of CBCT were slightly better than PA.[7,8] With unified and regular defect, particular settings, no soft coverage or other surrounding structures, the results should be interpreted with caution. On the other hand, Vidor et al. found that the agreements of PA were slightly better than CBCT, especially for inter-observer assessment with circumferential 0.125 mm peri-implant gap.[5] Meanwhile, the study

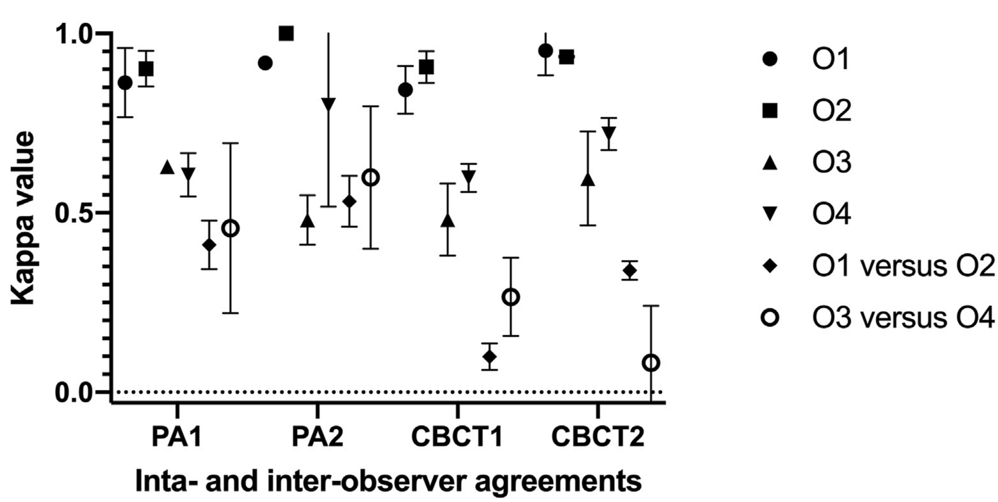

Figure 1 Intra- and inter-observer agreements of each observer for PA1 (original periapical radiography), PA2 (enhanced brightness/contrast periapical radiography), CBCT1 (selected axial and mesial-distal direction images) and CBCT2 (all data with software).

Table 4 Wilcoxon paired signed rank test for comparison between different categories.

Category	Median1	Median2	p value
Intra-observer PA versus CBCT	0.831 (PA)	0.774 (CBCT)	0.821
Inter-observer PA versus CBCT	0.471(PA)	0.192 (CBCT)	0.016*
Experienced observers versus inexperienced observers	0.883 (O1, O2)	0.567 (O3, O4)	$<$0.001*
PA1 versus PA2	0.629 (PA1)	0.670 (PA2)	0.278
CBCT1 versus CBCT2	0.561 (CBCT1)	0.688 (CBCT2)	0.176

PA, periapical radiography; CBCT, cone beam computed tomography; PA1, original periapical radiography; PA2, enhanced brightness/contrast periapical radiography; CBCT1, selected axial and mesial-distal direction images; CBCT2, all data with software.
*$p < 0.05$.

with 0.35 mm and 0.675 mm circumferential defects conducted by Dave et al. stated that PA showed significantly better intra-observer and inter-observer agreements than CBCT.[4] When peri-implant bone defects were small, artifact presence in CBCT would shallow bony defects, which could result in agreements of PA better than CBCT. Moreover, Pinheiro et al. reported that when peri-implant defects were chemically created, agreements of CBCT were better than PA.[6] In our study, different sizes of bone defects might be overdrilled. It is indicated that overdrilled defects show a dark band similar to that produced by metal artifacts of CBCT, which could go easily unnoticed compared to chemically created irregular defects.[9] Thus, in this respect, better agreements in PA could be expected in our clinical study.

Periapical films have been widely used in daily clinical practice and recommended as initial method for evaluating peri-implant bone status.[10,11] Although generally claimed as a standard method, diagnostic accuracy of PA is still under debate and information in agreements is still limited.[12,13] For example, a recent study suggested that marginal bone level changes in PA below 1 mm were likely due to human variation. Additionally, agreements within and between observers when assessing periapical radiographs were mostly 0.40—0.59, which was in agreement with our study. Moreover, brighter radiographs improved intra-observer agreement in determining peri-implant marginal bone level changes.[14] However, in our study, adjusted brightness/contrast radiographs brought little improvement in agreements. In fact, in our study, bone defects were mostly around apical portion of implants. As we all know, implant apical portion of periapical film is influenced by superimposed structures, variation in the trabecular pattern and defect's size, shape and density and the complexity of the surrounding normal anatomic features. Thus, assessment of apical portion bone of implants in PA is more complex than peri-implant marginal bone, which might explain the little improvement after film brightness/contrast adjustment.

It is somewhat surprising that the agreements of CBCT were poorer than PA, especially for inter-observer agreements. As many studies have proved, the diagnostic accuracy of CBCT was good.[7,8,15—18] For example, Golubovic et al. compared sagittal views of CBCT to histologic standard when evaluating advanced peri-implantitis defects and results showed CBCT could be regarded as an accurate diagnostic tool.[15] As a matter of fact, sagittal image could be least affected by artifacts with reduced gray value in peri-implant bone assessment. Another study proved that CBCT images correlated highly with physical measurements.[16] Meanwhile, the defect size in the their study was at least 1 mm and multiple sectional images were combined to diagnose defects. As pointed in a recent review, larger defect sizes exhibited a trend for better defect detection by CBCT.[17] On the other hand, Ritter et al. reported that PA and CBCT performed similar in assessing mesial and distal peri-implant marginal bone.[19] Meanwhile, other in vitro studies have stated that assessment of interproximal peri-implant defect width at implants was more accurate in PA in comparison to CBCT.[4,5,20] Smaller defect sizes were seen in those studies. Furthermore, in Steiger-Ronay's study, with only mesio-distal direction images of CBCT evaluated, PA was more accurate than CBCT.[20] Thus, variation in design of previous studies might explain the inconsistent diagnostic accuracy of CBCT.

As we all know, mesio-distal direction image of CBCT with implant artifacts is always present in the proximity of titanium implants with altered gray value, which could affect judgement prominently.[21] A pattern for the distribution of artifacts around titanium implants in CBCT was

Table 5 Cochran's Q-test with post hoc McNemar test results.

Observer	Cochran's Q-test		Post hoc McNemar test (p value)					
	Q value	p value	PA1PA2	PA1CBCT1	PA1CBCT2	PA2CBCT1	PA2CBCT2	CBCT1CBCT2
1	8.032	0.045*	0.317	0.157	0.046	0.096	0.025	1.000
2	4.435	0.218	0.564	0.167	1.000	0.285	0.564	0.166
3	16.948	0.001*	0.034	0.052	0.083	0.005**	0.001**	0.739
4	15.737	0.001*	0.180	0.257	0.018	0.034	0.002**	0.090

PA1, original periapical radiography; PA2, enhanced brightness/contrast periapical radiography; CBCT1, selected axial and mesial-distal direction images; CBCT2, all data with software.
*$p < 0.05$; **$p < 0.01$.

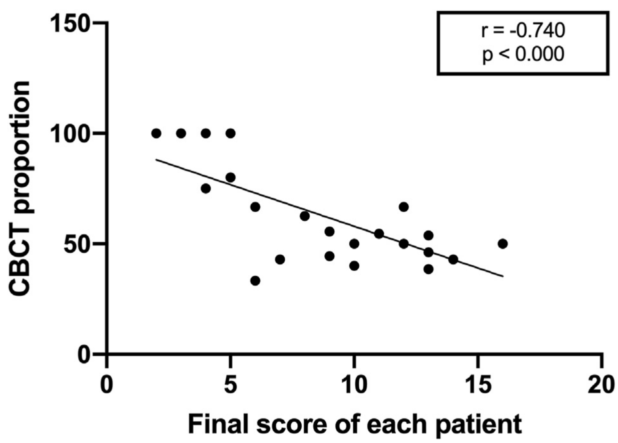

Figure 2 Spearman correlation between final score of each patient and proportion of CBCT.

detected. Regions with reduced gray values were located mesially and distally at molar, premolar and canine sites and at the mesio-buccal, disto-buccal, mesio-lingual and disto-lingual aspects of incisor sites.[22] On the contrary, regions with increased gray value were mostly located buccally and lingually, which would least affect peri-implant bone assessment. In our study, we were focusing on mesial-distal direction defect, which would be strongly affected by artifacts. Moreover, the presence of anatomical structures such as cranium, vertebral column, and soft tissue influence the gray value of the jawbones made the image in our study more complex to evaluate. Identification of artifact-affected regions is difficult, which could lead to false positive diagnosis of peri-implant bone defects. As the result of Spearman correlation showed, evaluations from CBCT were more likely to be considered as bone defect presence. Therefore, combined with inconsistent diagnostic accuracy and influence of artifacts, the poor agreements of CBCT could be explained.

Two methods were used in our study to assess peri-implant bone defect in CBCT. The first method was to assess only selected images, and the second method was to scroll all database of CBCT in each patient. Both methods were seen in previous studies, with the second method more frequently used. It is understandable that most studies have chosen the second method, which would allow observers to use any tools or any sectional images to view. In our study, we have compared the two methods. Our hypothesis was, by choosing the most representative image in CBCT1, diagnosis of bone defect could be facilitated. However, our results (agreements in CBCT2 were slightly better than CBCT1) have denied this hypothesis. The most possible explanation is that the sectional images in CBCT1 was selected by two researchers subjectively without knowing defect presence or not, thus complicated the assessment. On the contrary, by scrolling through all databases would give observers more information to diagnose.

Four observers participated in our study. Two experienced dentists with more than five years of experience in assessing radiographic image with implants and two graduate students major in implant dentistry. As for intra-observer agreements, results of experienced dentists were much better than graduate students. While for inter-observer agreements, there was small difference between experienced and inexperienced observers. Our result was different from Pelekos's study, which proved that the performance of PA film could be affected by experience, while accurate assessment of CBCT would not.[18] In fact, our study was a clinical study with irregular defect type and size. Moreover, in their study, the defect size was at least 1.7 mm. Other in vitro studies conducted by Pinheiro et al. found that kappa values for more experienced maxillofacial radiologists were better than maxillofacial surgeons,[9,23]

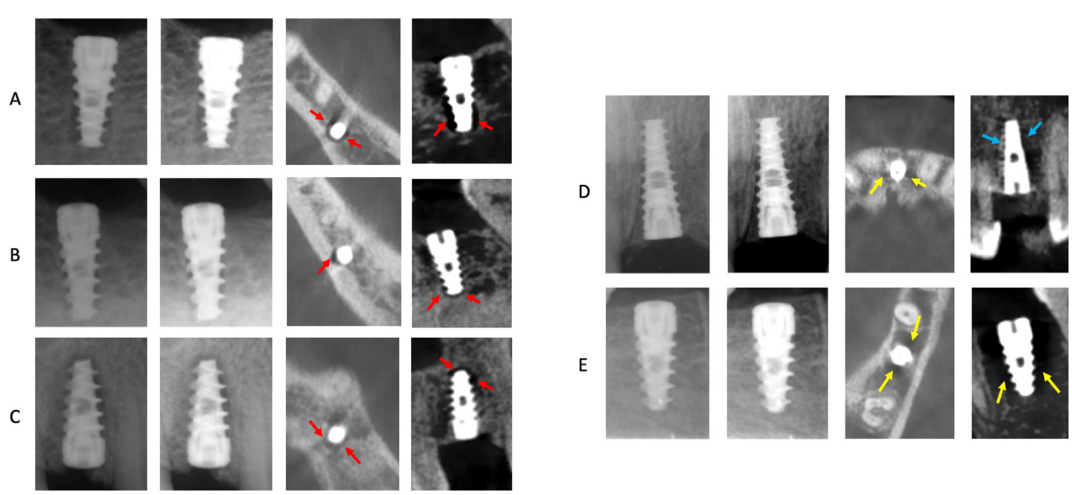

Figure 3 A, B, and C were the three patients with peri-implant bone defect identified by all 4 observers (red arrows point to the defects). D and E were the two patients with lowest scores (yellow arrows point to the artifacts with decreased gray value, while blue arrows point to the artifacts with increased gray value). Notice the upper incisor in D, artifacts with increased gray value were located at mesial and distal direction, while artifacts with decreased gray value was located at mesial-lingual and distal-lingual direction. Image sequence was PA1 (original periapical radiography), PA2 (enhanced brightness/contrast periapical radiography), CBCT axil view and CBCT mesial-distal view (from left to right).

which was in accordance with our study. Moreover, Cochran's Q value was smaller in observer 1 and 2 than in observer 3 and 4, which indicated smaller variation in diagnosis for experienced observers. It's noteworthy that inexperienced observers scored higher for CBCT and scored much less "3" (unsure) than experienced observers, which showed experienced dentists tend to be more careful with diagnosis in clinical situation.

There were only 3 patients whose diagnosis were always "1", which indicated confirmation of bone defects by all four observers both in PA and CBCT. As we can see from Fig. 3, the defect sizes in all 3 cases were not small. However, owing to undefined peri-implant defect in our study, the real defect size could not be acquired. With known 10 mm implants length placing in all 3 cases, we could only approximate the defect size subjectively. In fact, there were studies focusing on comparing defect sizes and evaluation results. There is a tendency that with defect size increasing, assessment scores would increase. The results of Kavadella's study showed that averaged score was 2.3 for the defect size 1.2 mm and 4.06 for the size 2.5 mm (same 5-point scale used in our study).[24] Dave et al. found that diagnostic accuracy was better for 0.675 mm defect than 0.35 mm space.[4] A recent study conducted by Kerkfeld et al. found bone defects that extending 400 um around implants could be reliably detected by using gray scale analysis in CBCT.[25] Therefore, in the future, with new technique applied, defect assessment around implants might be facilitated.

Nowadays, CBCT has been widely used in clinical practice. Meanwhile, when CBCT is needed, "As Low As Reasonably Achievable" principle must be applied. In our study, we have chosen the smallest FOV with 0.25 voxel size. However, different settings of CBCT are not only associated with dose reduction, but also important in detecting peri-implant defects. However, effects of those parameters such as voxel size, field of view and even metal artifact reduction appliance are still under debate.[16,26–28] In addition, different CBCT systems have also showed variation in detecting defects.[29,30] Continued research effort is urgently warranted to further explore these factors in diagnosing peri-implant defects with as low dose as possible in CBCT.

Although this study offers insights into radiographic agreements in clinical situation, several limitations should be considered. First, without knowing defect presence or not, we could not evaluate the diagnostic accuracy in PA and CBCT. Second, the geometric complexity of the bone defects is limited in the current study. Thus, the results might not be applicable to other clinical situation such as bone defects in peri-implantitis. Last but not the least, the use of only one imaging device for 2D and 3D imaging with only one parameter settings might reduce transferability of the results.

Within the limits of our study, agreements of PA are better than CBCT when detecting peri-implant bone defects, especially for inter-observer agreements. Experienced observers are more consistent in assessment than inexperienced ones. Individual variation in diagnosing defects around dental implants should be noted in clinical practice.

Declaration of competing interest

The authors have no conflict of interest to declare.

Acknowledgements

This work was financially supported by Science and Technology Commission of Shanghai Municipality (19411950100), China; and Multicenter Clinical Research Program of Shanghai Jiao Tong University School of Medicine (DLY201822), China.

References

1. Harris D, Horner K, Gröndahl K, et al. E.A.O. guidelines for the use of diagnostic imaging in implant dentistry 2011. A consensus workshop organized by the European Association for Osseointegration at the Medical University of Warsaw. *Clin Oral Implants Res* 2012;23:1243–53.
2. Jacobs R, Vranckx M, Vanderstuyft T, Quirynen M, Salmon B. CBCT vs other imaging modalities to assess peri-implant bone and diagnose complications: a systematic review. *Eur J Oral Implant* 2018;11:77–92.
3. Bohner LOL, Mukai E, Oderich E, et al. Comparative analysis of imaging techniques for diagnostic accuracy of peri-implant bone defects: a meta-analysis. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2017;124:432–40. e5.
4. Dave M, Davies J, Wilson R, Palmer R. A comparison of cone beam computed tomography and conventional periapical radiography at detecting peri-implant bone defects. *Clin Oral Implants Res* 2013;24:671–8.
5. Vidor MM, Liedke GS, Fontana MP, et al. Is cone beam computed tomography accurate for postoperative evaluation of implants? An in vitro study. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2017;124:500–5.
6. Pinheiro LR, Scarfe WC, de Oliveira Sales MA, Gaia BF, Cortes AR, Gusmão Paraiso Cavalcanti M. Effectiveness of periapical radiography versus cone beam computed tomography with different kilovoltage settings in the detection of chemically created peri-implant bone defects: an in vitro study. *Int J Oral Maxillofac Implants* 2017;32:741–50.
7. Schwindling FS, Hilgenfeld T, Weber D, Kosinski MA, Rammelsberg P, Tasaka A. In vitro diagnostic accuracy of low-dose CBCT for evaluation of peri-implant bone lesions. *Clin Oral Implants Res* 2019;30:1200–8.
8. Hilgenfeld T, Juerchott A, Deisenhofer UK, et al. Accuracy of cone-beam computed tomography, dental magnetic resonance imaging, and intraoral radiography for detecting peri-implant bone defects at single zirconia implants-an in vitro study. *Clin Oral Implants Res* 2018;29:922–30.
9. Pinheiro LR, Gaia BF, Oliveira de Sales MA, Umetsubo OS, Santos Junior O, Cavalcanti MG. Effect of field of view in the detection of chemically created peri-implant bone defects in bovine ribs using cone beam computed tomography: an in vitro study. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2015;120:69–77.
10. Kühl S, Zürcher S, Zitzmann NU, Filippi A, Payer M, Dagassan-Berndt D. Detection of peri-implant bone defects with different radiographic techniques - a human cadaver study. *Clin Oral Implants Res* 2016;27:529–34.
11. Raes F, Renckens L, Aps J, Cosyn J, De Bruyn H. Reliability of circumferential bone level assessment around single implants in healed ridges and extraction sockets using cone beam CT. *Clin Implant Dent Relat Res* 2013;15:661–72.

12. Cassetta M, Di Giorgio R, Barbato E. Are intraoral radiographs reliable in determining peri-implant marginal bone level changes? The correlation between open surgical measurements and peri-apical radiographs. *Int J Oral Maxillofac Surg* 2018;47:1358—64.
13. García-García M, Mir-Mari J, Benic GI, Figueiredo R, Valmaseda-Castellón E. Accuracy of periapical radiography in assessing bone level in implants affected by peri-implantitis: a cross-sectional study. *J Clin Periodontol* 2016;43:85—91.
14. Walton TR, Layton DM. Intra- and inter-examiner agreement when assessing radiographic implant bone levels: differences related to brightness, accuracy, participant demographics and implant characteristics. *Clin Oral Implants Res* 2018;29:756—71.
15. Golubovic V, Mihatovic I, Becker J, Schwarz F. Accuracy of cone-beam computed tomography to assess the configuration and extent of ligature-induced peri-implantitis defects. A pilot study. *Oral Maxillofac Surg* 2012;16:349—54.
16. Kamburoğlu K, Murat S, Kiliç C, et al. Accuracy of CBCT images in the assessment of buccal marginal alveolar peri-implant defects: effect of field of view. *Dentomaxillofac Radiol* 2014;43:20130332.
17. Pelekos G, Acharya A, Tonetti MS, Bornstein MM. Diagnostic performance of cone beam computed tomography in assessing peri-implant bone loss: a systematic review. *Clin Oral Implants Res* 2018;29:443—64.
18. Pelekos G, Tse JMN, Ho D, Tonetti MS. Defect morphology, bone thickness, exposure settings and examiner experience affect the diagnostic accuracy of standardized digital periapical radiographic images but not of cone beam computed tomography in the detection of peri-implant osseous defects: an in vitro study. *J Clin Periodontol* 2019;46:1294—302.
19. Ritter L, Elger MC, Rothamel D, et al. Accuracy of peri-implant bone evaluation using cone beam CT, digital intra-oral radiographs and histology. *Dentomaxillofac Radiol* 2014;43:20130088.
20. Steiger-Ronay V, Krcmaric Z, Schmidlin PR, Sahrmann P, Wiedemeier DB, Benic GI. Assessment of peri-implant defects at titanium and zirconium dioxide implants by means of periapical radiographs and cone beam computed tomography: an in-vitro examination. *Clin Oral Implants Res* 2018;29:1195—201.
21. Schulze RK, Berndt D, d'Hoedt B. On cone-beam computed tomography artifacts induced by titanium implants. *Clin Oral Implants Res* 2010;21:100—7.
22. Benic GI, Sancho-Puchades M, Jung RE, Deyhle H, Hämmerle CH. In vitro assessment of artifacts induced by titanium dental implants in cone beam computed tomography. *Clin Oral Implants Res* 2013;24:378—83.
23. Pinheiro LR, Scarfe WC, Augusto de Oliveira Sales M, Gaia BF, Cortes AR, Cavalcanti MG. Effect of cone-beam computed tomography field of view and acquisition frame on the detection of chemically simulated peri-implant bone loss in vitro. *J Periodontol* 2015;86:1159—65.
24. Kavadella A, Karayiannis A, Nicopoulou-Karayianni K. Detectability of experimental peri-implant cancellous bone lesions using conventional and direct digital radiography. *Aust Dent J* 2006;51:180—6.
25. Kerkfeld V, Meyer U. Higher resolution in cone beam computed tomography is accompanied by improved bone detection in peri-implant bone despite metal artifact presence. *Int J Oral Maxillofac Implants* 2018;33:1331—8.
26. Kamburoglu K, Kolsuz E, Murat S, Eren H, Yüksel S, Paksoy CS. Assessment of buccal marginal alveolar peri-implant and periodontal defects using a cone beam CT system with and without the application of metal artefact reduction mode. *Dentomaxillofac Radiol* 2013;42:20130176.
27. Vasconcelos KF, Codari M, Queiroz PM, et al. The performance of metal artifact reduction algorithms in cone beam computed tomography images considering the effects of materials, metal positions, and fields of view. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2019;127:71—6.
28. de-Azevedo-Vaz SL, Vasconcelos Kde F, Neves FS, Melo SL, Campos PS, Haiter-Neto F. Detection of periimplant fenestration and dehiscence with the use of two scan modes and the smallest voxel sizes of a cone-beam computed tomography device. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2013;115:121—7.
29. Eskandarloo A, Saati S, Ardakani MP, Jamalpour M, Gholi Mezerji NM, Akheshteh V. Diagnostic accuracy of three cone beam computed tomography systems and periapical radiography for detection of fenestration around dental implants. *Contemp Clin Dent* 2018;9:376—81.
30. Razavi T, Palmer RM, Davies J, Wilson R, Palmer PJ. Accuracy of measuring the cortical bone thickness adjacent to dental implants using cone beam computed tomography. *Clin Oral Implants Res* 2010;21:718—25.