Original Research Article

Interoperator Variations in Interpretation of Cone Beam Computed Tomography for Dental Implant Treatment Planning: A Retrospective Observational Study

Anas Abdul Khader*

Department of Periodontology and Implant Dentistry, College of Dentistry, Qassim University, Kingdom of Saudi Arabia
*Corresponding author email: **an.muhammed@qu.edu.sa**

International Archives of Integrated Medicine, Vol. 13, Issue 4, April, 2026.
Available online at http://iaimjournal.com/
ISSN: 2394-0026 (P) **ISSN: 2394-0034 (O)**
Received on: 2-4-2026 **Accepted on:** 10-4-2026
Source of support: Nil **Conflict of interest:** None declared.

DOI: **10.5281/zenodo.19822247**

How to cite this article: Anas Abdul Khader. Interoperator Variations in Interpretation of Cone Beam Computed Tomography for Dental Implant Treatment Planning: A Retrospective Observational Study. Int. Arch. Integr. Med., 2026; 13(4): 388-401.

Abstract

Background: Cone beam computed tomography (CBCT) is the gold standard for implant treatment planning, however variability in the interpretation may result in reduced diagnostic consistency and affect treatment outcomes.

Aim: To measure the diagnostic agreement for interpretation of CBCT scans for dental implant planning and determine the impact of operator experience on diagnostic agreement.

Materials and methods: In this retrospective observational study 100 anonymous CBCT scans of patients needing implant treatment planning were assessed by two groups of operators: Group A: clinicians with >=10 years of experience and Group B: clinicians with <=5 years of experience. The operators assessed bone height, width, Hounsfield unit (HU) values, bone density (Misch D1-D4) and implant feasibility including recommended surgical modification. The reliability of operator agreement was assessed using the intraclass correlation coefficient (ICC), Cohen's kappa, t-test and Bland-Altman analysis.

Results: There was an excellent agreement amongst operators for linear measurements (ICC = height = 0.996, width = 0.989). Density measurements showed an excellent correlation (ICC = 0.984), but a statistically significant systematic difference (mean difference = 16.31 HU, p = 0.001) was observed, with higher values in experienced operators. Bone density determination showed substantial agreement ($\kappa = 0.793$; weighted $\kappa = 0.830$), with all 12 disagreements occurring between consecutive categories, most commonly at the D2/D3 interface (58.3%). The assessment of implant feasibility was found to be clinically relevant. Group A rated 90% of the sites feasible compared to 80% by Group B (k = 0.668). The 10 dissenting feasibility assessments (feasibility vs. non-feasibility) were all recommended by the experienced clinicians. Most disagreements (6/14) were in the Posterior Maxilla due to sinus-related complexities.

Conclusions: CBCT linear measurements showed excellent interoperator agreement, independently of the clinical experience, however, bone density assessment and its feasibility for dental implant placement were based on operator's experience. Experienced operators showed larger armamentarium, thus more positive feasibility. Guidelines for how to interpret and dedicated training in more advanced surgical options may reduce the variability due to experience and increase the consistency of planning.

Key words

Cone beam computed tomography; dental implants; interoperator agreement; treatment planning; bone density; Hounsfield units; clinical experience; intraclass correlation coefficient.

Introduction

Cone beam computed tomography (CBCT) has revolutionised the planning of dental implants by enabling three-dimensional volumetric imaging of the jaw bones, hence providing the opportunity to assess bone shape, quality and proximity of critical anatomical structures [1, 2]. CBCT allows multiplanar reconstructions that provide accurate measurements of bone size, location of the inferior alveolar nerve canal, shape of the maxillary sinus and quality of bone, all of which have been shown to affect the outcome of dental implants [3, 4].

For all its technical benefits, CBCT is still a subjective examination. The diagnostic accuracy of CBCT is reliant upon the clinician's ability to accurately interpret anatomical landmarks, measure anatomical bone indices, assess the density of boneand integrate these factors into a meaningful treatment plan [5]. As a result, variability in the interpretation of CBCTs between operators is of great concern and differential diagnostic findings may lead to sub-optimal choice of implants, surgical approaches and treatment outcomes [6, 7].

Variability in the interpretation of radiographs has been attributed to a number of factors. The most important factors affecting the accuracy and reliability of diagnosis are clinical experience and training [8, 9]. Some studies have demonstrated that the operator experience plays a significant role in interobserver agreement on the diagnosis of peri-implant bone defects, and the agreement of the experienced clinicians is higher than less experienced ones [10]. Likewise, Pelekos, et al. reported the difficulties in the diagnosis due to the presence of artefacts and complex anatomical structures, which may impact more on operators with less experience [11].

CBCT interpretation is difficult, particularly when assessing bone density. The bone density classification system described by Misch (D1 - D4), which reflects the cortical thickness and density is commonly used for implant planning [3]. However, there is an opinion that assigning Hounsfield unit (HU) values to discrete bone density groups includes subjective cut-offs which may vary between the operators particularly at the interface of the groups [12, 13]. Moreover, the HU values of the CBCT are influenced by scan settings, FOV and reconstruction algorithms, further complicating interpretation [14, 15].

In addition to measurement accuracy, clinical judgment plays a role in the diagnosis of CBCT images (decision making on implant suitability and choice of surgical alteration) and may therefore introduce another source of interoperator variability. An operator with greater experience and a greater repertoire of surgical skills may identify treatment possibilities for sites deemed unsuitable for implant placement by less experienced operators [16, 17]. This experience-dependent difference in decisions has significant clinical implications for patient care as it can result in either a failure to treat or a mistreatment of a condition, based upon the experience of the operator.

The variability of CBCT interpretation has been studied primarily for certain diagnostic purposes, such as peri-implant bone defect [10, 18], root fracture [19], and endodontics [20]. But only a few studies have been subjected to detailed analyses of the entire range of interoperator agreement for treatment planning with dental implants, including linear measurements, bone density classification, anatomical structure identification and their feasibility of treatment all at once.

The purpose of the current study is to determine the nature of the interoperator variability associated with the interpretation of CBCTs for implant treatment planning. More specifically, to: (1) determine interoperator agreement in linear bone height and HU value determination (2) determine agreement in classification of bone volume density using the Misch system (3) determine agreement in implant feasibility assessment and surgical modification recommendations; and (4) determine the influence of experience for all aspects of CBCT assessment. These variations must be understood in order to establish standard operating procedures that aim to improve diagnostic consistency and improve implant treatment outcomes.

Materials and methods

Study Design and Ethics Approval

This is a retrospective and observational study conducted at the Department of Oral Radiology of the College of Dentistry, Qassim University, Saudi Arabia. Ethical approval was obtained from the Qassim University Institutional Review Board (Reference: 25-37-18). Due to the retrospective nature of the study, and the use of de-identified imaging data, the requirement for informed consent was waived. The study was done according to the Declaration of Helsinki and Institutional ethical guidelines.

Sample Size Determination

The calculation of the sample size was based on the interobserver agreement data reported by Zhang et al.[10] in which the values of Cohen's kappa ranged between 0.192 and 0.883 for different levels of operator experience. Using the formula n = Z2P(1 - P)/D2, where Z = 1.96 (confidence level considered: 95%), P = 0.7 (estimated value for the proportion of agreement) and D = 0.1 (desired precision), a minimum sample size of 81 scans was calculated. In order to account for potential variability, 100 CBCT scans were included in the final analysis.

Sample Selection

CBCT images were retrospectively accessed from the Departmental imaging database. Inclusion criteria were: (1) scans taken from patients aged 18-75 years who had been referred for dental implant treatment planning; (2) scans for which no previous annotations or digital markings had been made; and (3) scans that were of diagnostic quality with adequate field of view that spanned the region of interest. Exclusion criteria were: (1) scans with significant image artefacts compromising diagnostic quality; (2) patients with systemic conditions known to affect bone density (i.e. uncontrolled diabetes, osteoporosis, bisphosphonate therapy); and (3) scans with incomplete or truncated fields of view.

Operator Groups

Two groups of operators were recruited with the help of clinical experience. Group A consisted of clinicians who had 10 or more years of experience in CBCT interpretation and implant treatment planning. Group B included groups of clinicians with 5 or less years of experience. All the operators had postgraduate qualifications in dentistry, and were actively involved with implant treatment planning. Operators were blinded from each other's assessments throughout the study.

CBCT interpretation protocol

Each operator independently assessed all 100 CBCT scans using a standardised form of assessment. The following parameters were evaluated: (1) available bone height (mm), measured from the alveolar crest to the relevant anatomic boundary (mandibular canal, maxillary sinus floor or nasal fossa floor); (2) available bone width (mm), assessed at the narrowest buccolingual dimension of the alveolar ridge, typically measured 2–3 mm apical to the alveolar crest for accurate dental implant planning; (3) Hounsfield unit (HU) value at the proposed implant site; (4) bone density classification according to the Misch system (D1: dense cortical bone; D2: thick cortical bone with coarse trabecular bone; D3: thin cortical bone with fine trabecular bone; D4: very thin cortical bone with low-density, fine trabecular bone).Scans were distributed in four anatomical regions: Posterior Maxilla (n = 30), Anterior Maxilla (n = 26), Anterior Mandible (n = 23) and Posterior Mandible (n = 21). All assessments were conducted on calibrated high-resolution monitors under standardised viewing conditions.

Statistical Analysis

Descriptive statistics (mean, standard deviation, median, range) were calculated for all continuous measurement parameters. Interoperator agreement of the continuous variables (height, width, HU values) was evaluated with intraclass correlation coefficients (ICC, two-way random effects model, single measures), Pearson correlation, paired t test, and Bland-Altman analysis with 95% limits of agreement. (LoA) For categorical variables, Cohen's kappa and linearly weighted kappa were determined for bone density classification and Cohen's kappa was determined for implant feasibility assessment. Region-specific agreement statistics were calculated in order to point out anatomical regions with the highest interpretive variability. Statistical Significance was defined as $p < 0.05$. All statistical analyses were conducted in the software package, IBM Statistical Package for Social Sciences (SPSS) version 28.0 (IBM Corp., Armonk, NY, USA).

Results

Sample Characteristics

A total of 100 CBCT scans were independently evaluated by 2 groups of operators - Group A, which includes clinicians with 10 or more years of experience, and Group B, which includes clinicians with 5 or fewer years of experience. The distribution of the scans in the four anatomical regions was Posterior Maxilla (n=30, 30%), Anterior Maxilla (n=26, 26%), Anterior Mandible (n=23, 23%), and Posterior Mandible (n=21, 21%). Operators measured the available bone height (mm), width (mm), Hounsfield unit (HU) values, bone density classification (D1-D4), and implant feasibility assessment, including identification of relevant anatomical structures and recommended surgical modifications.

Descriptive Statistics of Linear and Density measurements

Table - 1 shows the descriptive statistics of three continuous measurement parameters for both operator groups. The mean available bone height was almost identical in both groups (Group A: 11.08 +- 4.20 mm; Group B: 11.08 +- 4.25 mm) matching medians of 11.20 mm. Width measurements indicated similar concordance (Group A: 6.21 +- 1.93 mm; Group B: 6.17 +- 1.91 mm). Hounsfield unit values also showed a slight systematic difference with Group A recording slightly higher mean values (747.00 +- 289.65) than Group B (730.69 +- 290.99).

Interoperator Agreement of Continuous Measurements

Interoperator reliability for continuous measurements was evaluated by computing intraclass correlation coefficients (ICC), Pearson correlation coefficients, paired t-tests, and Bland-

Altman analysis (**Table - 2**). All three continuous parameters showed excellent interoperator agreement. Height measurements were the most reliable measurements (ICC = 0.996, r = 0.996), followed by width (ICC = 0.989, r = 0.989) and HU values (ICC = 0.984, r = 0.985).

Table – 1: Descriptive statistics of continuous measurement parameters by operator group.

Parameter	**Group A Mean±SD**	**Median**	**Range**	**Group B Mean±SD**	**Median**	**Range**
Height (mm)	11.08±4.20	11.20	1.5–18.1	11.08±4.25	11.20	1.9–18.4
Width (mm)	6.21±1.93	6.25	2.3–10.4	6.17±1.91	6.30	2.5–10.3
HU Value	747.00±289.65	762.00	203–1397	730.69±290.99	726.00	187–1383

Table – 2: Interoperator agreement statistics for continuous CBCT measurements.

Parameter	**Pearson r**	**ICC (2,1)**	**Mean Diff ± SD**	**Paired t**	**95% LoA**	**p-value**
Height (mm)	0.996	0.996	0.001 ± 0.376	0.027	−0.736 to 0.738	0.979
Width (mm)	0.989	0.989	0.035 ± 0.290	1.207	−0.533 to 0.603	0.228
HU Value	0.985	0.984	16.31 ± 50.15	3.252	−82.0 to 114.6	0.001*

** Statistically significant at p < 0.05.*

Table – 3: Bone density classification (D1–D4) distribution by anatomical region and operator group.

Region	**D1 (A)**	**D1 (B)**	**D2 (A)**	**D2 (B)**	**D3 (A)**	**D3 (B)**	**D4 (A)**	**D4 (B)**
Ant. Mand. (23)	6 (26.1%)	3 (13.0%)	16 (69.6%)	20 (87.0%)	1 (4.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Post. Mand. (21)	0 (0.0%)	0 (0.0%)	2 (9.5%)	3 (14.3%)	19 (90.5%)	18 (85.7%)	0 (0.0%)	0 (0.0%)
Ant. Max. (26)	0 (0.0%)	0 (0.0%)	9 (34.6%)	6 (23.1%)	17 (65.4%)	20 (76.9%)	0 (0.0%)	0 (0.0%)
Post. Max. (30)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	21 (70.0%)	19 (63.3%)	9 (30.0%)	11 (36.7%)

Paired t-tests did not show any statistically significant differences between operator groups in height (mean difference = 0.001 mm, p = 0.979) or width measurements (mean difference = 0.035 mm, p = 0.228). However, there was a small but statistically significant interoperator difference in HU (mean difference = 16.31 HU, t = 3.252, p = 0.001) with Group A always reporting higher values.

Bland-Altman analysis confirmed narrow limits of agreement for height (-0.736 to 0.738 mm) and width (-0.533 to 0.603 mm), and thus discrepancies between operators for linear measurements were clinically negligible. The 95% limits of agreement for HU values were broader (-81.98 to 114.60 HU), indicating higher variation of density assessment among different operators with varying levels of experience.

Classification of Bone Density according to Anatomical Region

The distribution of the bone density classifications by anatomical locations is presented in **Table - 3**. Both operator groups showed consistent region-specific densities in the following groups: the Anterior Mandible was mostly in the D1-D2 group, the Posterior

Mandible and Anterior Maxilla were in the D2-D3 group, and the Posterior Maxilla was in the D3-D4 group.

A significant interoperator pattern was observed in which Group A classified a higher proportion of Anterior Mandible sites as D1 (26.1%) than did Group B (13.0%), but that Group B assigned more sites to the D2 category (87.0% vs. 69.6%). Similarly, amongst the Posterior Maxilla, more sites were classified as D4 in Group B (36.7%) than in Group A (30.0%). This pattern suggests that the experienced clinicians (Group A) were more prone to apply higher-density classifications in the boundary zones.

Table – 4: Hounsfield unit values by bone density classification and operator group.

Bone Type	**n (A)**	**Mean±SD (A)**	**Range (A)**	**n (B)**	**Mean±SD (B)**	**Range (B)**
D1	6	1296.2	1252–1397	3	1361.0	1347–1383
D2	27	1015.8	839–1241	29	1027.4	780–1242
D3	58	639.9	372–845	57	637.5	359–845
D4	9	264.3	203–312	11	259.5	187–346

Table – 5: Cross-tabulation of bone density classification between Group A and Group B.

Group A \ Group B	**D1**	**D2**	**D3**	**D4**	**Total**
D1	3	3	0	0	6
D2	0	23	4	0	27
D3	0	3	53	2	58
D4	0	0	0	9	9

Ranges of HU Values According to Bone Density Classification

Table - 4 shows the ranges of HU values that correspond to each of the bone density classifications assigned by each of the groups of operators. Both groups showed distinct separation of HU ranges between bone types, with significant overlap of the HU ranges assigned. The D1 classification had HU values greater than 1250 for both populations, D2 was between 780-1242, D3 was 359-845; D4 was 187-346.

Of particular note, the ranges for HU values of adjacent bone types overlapped (e.g. D2 upper range of 1241-1242 vs. D1 lower range of 1252-1347; D3 upper range of 845 vs. D2 lower range of 780-839), supporting that the boundaries between bone density categories are continuous transition zones, rather than discrete values. Group B had broader HU ranges for D4 classification (187-346) than for Group A (203-312), indicating that less experienced operators are more likely to use more liberal criteria for the lowest density category.

Interoperator Agreement for Bone Density Classification

Bone type classification concordance was 88% (88/100 cases). Cohen’s kappa was 0.793 (substantial agreement) and the linearly weighted kappa was 0.830. The cross-tabulation of classifications is shown in **Table - 5**.

All 12 disagreements were between adjacent bone density categories (that is 100% were off-by-one disagreements). The most common area of disagreement was D2/D3 (7/12, 58.3%), followed by D1/D2 (3/12, 25.0%) and D3/D4 (2/12, 16.7%). Bone type agreement was best observed in the Posterior Maxilla (93.3%) and lesser in the Anterior Mandible (82.6%).

Table – 6: Distribution of implant feasibility assessment categories by operator group (N = 100).

Feasibility Category	**Group A n (%)**	**Group B n (%)**	**Diff**	**Agreement**
Ideal dimension	54 (54.0%)	50 (50.0%)	+4	κ = 0.668
Horizontal augmentation	8 (8.0%)	6 (6.0%)	+2	
Short implant	7 (7.0%)	9 (9.0%)	−2	
Indirect sinus lift	6 (6.0%)	6 (6.0%)	0	
Lateral window sinus lift	6 (6.0%)	4 (4.0%)	+2	
Ridge expansion	4 (4.0%)	4 (4.0%)	0	
Ridge expansion + augment.	2 (2.0%)	1 (1.0%)	+1	
GBR augmentation	1 (1.0%)	0 (0.0%)	+1	
Vert. augment./nerve repos.	1 (1.0%)	0 (0.0%)	+1	
Pterygoid/zygomatic implants	1 (1.0%)	0 (0.0%)	+1	
Not feasible	10 (10.0%)	20 (20.0%)	−10	
Total (Feasible)	90 (90.0%)	80 (80.0%)	+10	

Anatomical Structure Identification and Implantation Feasibility Study

The distribution of implant feasibility categories as judged by each group of operators is shown in **Table - 6**. A total of 11 different feasibility classes were found, which included different anatomical considerations and surgical modification recommendations. The most common category assigned was "Ideal dimension" (Group A: 54%; Group B: 50%) indicating sites with sufficient bone for standard implant placement without surgical modification. The most clinically significant finding was the divergence in the binary feasibility determination. Group A (experienced clinicians) judged 90 sites as feasible for implant placement (with or without modification) and Group B (less experienced clinicians) judged only 80 sites as feasible ($p < 0.05$, k = 0.668). Group A found more examples of surgical modifications that may make compromised sites acceptable for implant placement, including GBR augmentation, vertical augmentation accompanied by nerve repositioning, and pterygoid/zygomatic implants - none of which were suggested in Group B.

Considerations: Regional Anatomical

Anterior Mandible (n = 23): Ideal dimension sites were mainly recognised by both groups (Group A: 73.9%; Group B: 65.2%). Ridge expansion was recommended the same (17.4% each). Group A found one site in need of GBR prior to implant placement whereas Group B found this site not feasible. The inferior alveolar nerve was not a dominant consideration in this region, consistent with its anterior location.

Posterior Mandible (n = 21): This area had the greatest percentage of recommended short implants (Group A: 33.3%; Group B: 38.1%) based on low available bone height because of the proximity to the inferior alveolar nerve canal. Group A was the only group who identified one case where vertical augmentation or repositioning of the nerve was required, a difficult surgical technique that was not recommended by Group B, who instead recorded this site as not feasible. In both groups the not-feasible rate was low (A: 4.8%; B: 9.5%).

Anterior Maxilla (n = 26): Horizontal augmentation of the bone was by far the most common anatomical issue suggested in Group A (30.8%) and B (23.1%) of instances. This is due to the usual occurrence of labial bone deficiency in the anterior maxillary region. Both groups agreed on ideal dimension sites at the same rate (57.7%). Group B classified more sites as not feasible (19.2% vs. 11.5%), which may indicate that less experienced clinicians are underestimating the potential of horizontal augmentation procedures.

Posterior Maxilla (n = 30): This area showed the greatest level of anatomical consideration and the highest level of interoperator divergence. Assessments related to sinus were high, and both groups recommended indirect sinus lift at similar rates (Group A: 20.0%; Group B: 20.0%). However, Group A recommended lateral window sinus lift more often (20.0% vs. 13.3%), and described the pterygoid/zygomatic implant options (3.3%) that were not considered by Group B. The not feasible rate differed most markedly in this region (Group A: 16.7%; Group B: 33.3%), an interpersonal difference of two-fold due to the experience of Group A in advanced surgical approaches to the proximity of the maxillary sinus.

Table – 7: Detailed analysis of all interoperator feasibility disagreement cases.

Scan	Region	Group A Assessment	Group B Assessment	Pattern
3	Ant. Mandible	Ridge expansion	Not feasible	A→Feasible
6	Ant. Mandible	GBR augmentation	Not feasible	A→Feasible
56	Ant. Mandible	Ideal dimension	Short implant	Approach diff
92	Ant. Mandible	Ideal dimension	Ridge expansion	Approach diff
12	Post. Mandible	Vert. augment./nerve repos.	Not feasible	A→Feasible
77	Post. Mandible	Ideal dimension	Short implant	Approach diff
16	Ant. Maxilla	Horizontal augmentation	Not feasible	A→Feasible
20	Ant. Maxilla	Horizontal augmentation	Not feasible	A→Feasible
22	Post. Maxilla	Lateral window sinus lift	Indirect sinus lift	Approach diff
25	Post. Maxilla	Lateral window sinus lift	Not feasible	A→Feasible
27	Post. Maxilla	Ridge expan. + augment.	Not feasible	A→Feasible
28	Post. Maxilla	Pterygoid/zygomatic	Not feasible	A→Feasible
29	Post. Maxilla	Ideal dimension	Not feasible	A→Feasible
30	Post. Maxilla	Indirect sinus lift	Not feasible	A→Feasible

Detailed Analysis of Feasibility Disagreement Cases

Fourteen cases (14%) presented disagreement on feasibility assessment between the two groups of operators. **Table - 7** shows each case of disagreement with the specific assessments and nature of discordance.

A unidirectional pattern was evident whereby in all of the 10 cases in which the disagreement was about a feasibility versus non-feasibility determination, Group A (experienced clinicians) judged the site as feasible with surgical modification and Group B judged the site as not feasible. No cases were seen where Group B found a site to be feasible where Group A found it not feasible ($p < 0.01$ by sign test). The remaining 4 disagreement cases involved differences in the recommended surgical approach but not fundamental divergence in feasibility.

The disagreement cases were distributed in all the regions: Anterior Mandible (4 cases), Posterior Maxilla (6 cases), Anterior Maxilla (2 cases), Posterior Mandible (2 cases). The Posterior Maxilla was the major contributor of disagreements with 5 out of 6 disagreements because Group A accepted manageability by sinus augmentation, ridge expansion, or zygomatic implant techniques for the sites which Group B considered not feasible.

Table – 8: Region-wise interoperator agreement statistics.

Region (n)	ICC Ht	ICC Wd	ICC HU	Bone %	Feas %	Mean |ΔHU|	Max |ΔHU|
Ant. Mandible (23)	0.992	0.982	0.911	82.6	82.6	54.6	190
Post. Mandible (21)	0.990	0.988	0.906	85.7	90.5	40.0	78
Ant. Maxilla (26)	0.981	0.990	0.905	88.5	92.3	43.8	88
Post. Maxilla (30)	0.989	0.983	0.937	93.3	80.0	39.6	109

Effect of Operator Experience on Anatomical Interpretation

The region-wise analysis (**Table - 8**) found that the effect of operator experience on different aspects of CBCT interpretation was differential.

Experienced clinicians (Group A) demonstrated a broader spectrum of surgical planning decisions, utilizing 11 distinct feasibility categories compared to 8 in Group B. Advanced procedures - such as guided bone regeneration (GBR), vertical augmentation with nerve repositioning, and pterygoid or zygomatic implant placement - were predominantly proposed by Group A, indicating a higher procedural confidence and familiarity with complex interventions. Experienced clinicians appeared to apply a more stringent threshold for deeming sites non-feasible.

Bone density assessment demonstrated a systematic experience-related variation, with Group A assigning significantly higher Hounsfield Unit (HU) values (mean difference = +16.31 HU, p = 0.001). This tendency was particularly evident at the D1/D2 interface, where experienced clinicians were more likely to classify borderline regions as higher-density bone, especially in the anterior mandible.

Sources of Variability of CBCT Interpretation Bone Density Assessment (HU Values)

Hounsfield unit measurement was the continuous parameter which showed the most variability from operator to operator. While the overall ICC (0.984) showed a great degree of agreement, regional ICCs varied from 0.905 (Anterior Maxilla) to 0.937 (Posterior Maxilla). The statistically significant systematic difference (mean = 16.31 HU, p = 0.001) indicates that the strategies of ROI placement or selection criteria of slicing are affected by operator experience. The 95% limits of agreement (-81.98 to 114.60 HU) are approximately 197 HU, which is clinically relevant given that it cuts across bone density category boundaries.

Thresholds for Classification of Bone Density

All of the 12 disagreements in bone type were at transitional HU-zones between adjacent categories. The following D2/D3 boundary (around 780-900 HU) was the most common source of discordance (58.3%), followed by the D1/D2 boundary around 1210-1290 HU (25.0%) and the D3/D4 boundary around 330-420 HU (16.7%). The findings suggest that there are zones of interpretive ambiguity in the Misch bone density classification system.

Anatomy of Complexity and Surgical Decision Making

The region of the greatest anatomic complexity and the largest divergence between operators was the Posterior Maxilla. The presence of the maxillary sinus necessitated decision-making by operators regarding residual bone height, sinus membrane integrity, and the suitability of various sinus augmentation techniques - decisions that varied widely depending on experience level. In contrast, the Anterior Maxilla had the highest feasibility agreement (92.3%) despite having to assess the horizontal augmentation, which suggests decisions about width are more straightforward than decisions about sinuses.

Surgical Knowledge and Treatment Planning Variability

The presence of the most significant variance was the differential surgical knowledge between operator groups; Group A always deeming feasible and Group B always deeming not feasible. This 10-case unidirectional disagreement pattern indicates a systematic experience-related bias and not random measurement error. Less experienced clinicians seem to make more conservative judgements about feasibility, which may be because they are unfamiliar with advanced surgical procedures such as nerve repositioning, pterygoid implants, and complicated sinus augmentation procedures.

Discussion

This study is a comprehensive evaluation of the interoperator variations in CBCT interpretation in the treatment planning of dental implants, which includes linear measurements, bone density evaluation, and clinical decision making. The results show that while quantitative measurements have excellent interoperator reliability, qualitative assessments and clinical judgements are strongly affected by operator experience, with important implications for the standardisation of the diagnostics, as well as clinical education.

The excellent interoperator agreement witnessed for linear measurements (ICC > 0.98 for both height and width) are consistent with previous reports. Fokas, et al. [21] found ICC results above 0.95 for linear measures derived from CBCT in the mandibular canal region and Ferrare, et al. [22] demonstrated similar reliability for those measures in the maxilla. The fact that the mean bone height values between groups were virtually identical (11.08 mm for both groups) and that the BMD narrow Bland-Altman limits of agreement were narrow (between - 0.736 and 0.738 mm for height; between - 0.533 and 0.603 mm for width) confirms the findings of high reproducibility with relatively low experience-related bias of linear measurements on CBCT. This finding supports the reliability of CBCT in dimensional assessment in routine clinical practice regardless of the experience level of the operator.

In contrast, bone density assessment became the parameter that was most prone to errors arising from operator differences. Although the overall ICC for HU values (0.984) show excellent agreement, the significant (p=0.001) systematic difference (mean = 16.31 HU) indicates that there is an experience-related bias in density measurement. This systematic offset probably reflects differences in region-of-interest (ROI) placement strategies, in which experienced clinicians may select more representative cortical and trabecular bone regions and less experienced operators may include adjacent soft tissue or cortical boundaries in their measurements14,23. The 95% limits of agreement with an approximate range of approximately 197 HU are clinically relevant in that this range straddles well established bone density category thresholds and may lead to different classification outcomes for a given site.

The fact that the discrepancy in bone density classification was confined to adjacent categories in all the cases is interesting and clinically significant. This pattern suggests that disagreements are not random but are a manifestation of a systematic ambiguity at transitional zones between categories of bone density. The D2/D3 boundary (around 780-900 HU) was the most common cause for disagreement, 58.3% of disagreements. This is clinically significant as the difference between the D2 and D3 bone often affects the choice of implant protocol; including drilling sequences, installation torque recommendations, and loading timelines [23, 24]. The categorical nature of the Misch classification system, though clinically practical, by its very nature leaves zones of interpretive ambiguity that may not adequately reflect the continuum of bone density variation [25].

The region-specific analysis of bone density classification provides more nuance. In the

Anterior Mandible, where dense cortical bone predominates, Group A described 26.1% of sites as D1 but only 13.0% by Group B. This regional pattern indicates that experienced operators may be better calibrated to identify the densest bone categories, whereas less experienced operators default to intermediate classifications. The result that Group B assigned wider HU ranges for D4 bone (187-346 vs. 203-312) is also consistent with the idea that less experienced clinicians use more relaxed and potentially less precise criteria for classification.

The most clinically significant finding from this study was the marked divergence in the assessment of implant feasibility between groups of operators. Experienced clinicians considered 90% of sites feasible, compared to 80% by less experienced operators, a clinically meaningful difference of 10 percentage points that could have an impact on treatment access in one in ten patients. Moreover, the fact that all of the disagreements about feasibility were unidirectional (that is, experienced operators came to recognize surgical solutions for sites that were considered unfeasible by less experienced colleagues) speaks against random measurement errors and argues for a systematic experience-related bias in clinical decision-making.

This divergence is due to the wider range of surgical feats demonstrated by experienced surgeons, who used 11 different feasibility categories compared with 8 by less experienced operators. Advanced techniques such as guided bone regeneration, vertical augmentation associated with nerve repositioning, and pterygoid/zygomatic implant position were only recommended by Group A. This finding is consistent with the concept of the "expertise paradox" in diagnostic imaging, in which advanced knowledge allows clinicians to see things (i.e., possibilities) that are invisible to less experienced clinicians [26, 27]. The clinical implication is significant: It is possible for patients to be unnecessarily denied implant treatment or referred for more conservative alternatives when such advanced solutions exist when they are evaluated by less experienced clinicians.

The Posterior Maxilla had the most interoperator disagreement in feasibility assessment (83.3% vs. 66.7%), accounting for 6 out of 14 disagreement cases. Five of these concerned sites which were deemed to be manageable by Group A through sinus augmentation technique or zygomatic implant technique which Group B deemed not feasible. This finding represents the complexity of the anatomical considerations of this region, such as the proximity of the maxillary sinus, assessment of the residual bone height, and choice of one of the different sinus augmentation approaches. Conversely, the Anterior Maxilla had the highest agreement in feasibility (92.3%) indicating that the horizontal evaluation of bone is a more simple interpretive task than the decision-making process relating to the sinuses. These results indicate that specific training in posterior maxillary anatomy and sinus augmentation technique may achieve maximum improvement in interoperator consistency.

The Anterior Mandible had the highest absolute HU discrepancy (mean = 54.6 HU, max = 190 HU), and lowest bone density classification agreement (82.6%). This region, which is characterised by dense cortical bone, poses special problems for density estimation since minute changes in the position of the ROI within the cross-section can result in very different HU values due to the transition between dense cortical and less dense trabecular compartments [28]. This finding highlights the importance of having standardised ROI placement protocols for various anatomical regions.

There are a few limitations that should be acknowledged. First, the retrospective design prevented control of the scanning parameters and the selection of patients. Second, the dichotomous categorization of operators as highly or less experienced (10 years of or less than 5 years) may not reflect the continuum of expertise development. Third, the study was done at one institution and may not be generalizable. Fourth, the study did not measure intraoperator reliability, which would help provide more context for interpreting interoperator variability. Finally, the correlation between HU values obtained by CBCT and true bone density are affected by scanner-specific factors, and the results may not be directly extrapolated between different CBCT systems [14, 15].

Future research should investigate the effectiveness of structured training programmes and standardised protocols for interpretation of data in reducing interoperator variability. The development of artificial intelligence-driven CBCT interpretation tools may prove to be a promising way to improve diagnostic consistency, especially for less experienced clinicians [29, 30]. Longitudinal studies examining the progression of interpretive accuracy at various levels of clinical training would add more information to the development of curricula in implant education. Additionally, multi-centre studies involving the use of different CBCT systems would increase the generalisability of results on interoperator agreement.

Conclusion

This study has shown that interoperator agreement between interpretation of CBCT for dental implant treatment planning varies considerably with different types of assessment parameters. Linear measurements of the height and width of bone show an excellent reliability (ICC > 0.98) regardless of experience in use, confirming the robustness of CBCT for dimensional assessment. However, the evaluation of bone density and implant feasibility determination is greatly affected by clinical experience, as clinical experience tends to record systematically higher HU values, placing borderline sites in higher-density categories, and identifying surgical solutions for sites deemed non-feasible by less experienced clinicians. The one-way communication that occurs in all 10 of the feasibility vs. non-feasibility disagreements shows the importance of clinical experience and surgical knowledge in CBCT-based treatment planning. Experienced clinicians used 11 different types of feasibility categories compared to 8 by less experienced operators, with such advanced techniques as GBR, repositioning nerves, and pterygoid/zygomatic implants recommended only by the experienced group. The Posterior Maxilla was the region that showed the most interoperator divergence, highlighting the specific complexity of sinus related surgical decision making.

These findings highlight the importance of standardised CBCT interpretation protocols, structured training in bone density evaluation and ROI placement as well as thorough education in advanced surgical techniques to reduce experience-related variability, and provide equitable treatment planning outcomes for patients.

References

1. Scarfe WC, Farman AG. What is cone-beam CT and how does it work? Dent Clin North Am. 2008;52(4):707-730.
2. Bornstein MM, Scarfe WC, Vaughn VM, Arcúris J. Cone beam computed tomography in implant dentistry: a systematic review focusing on guidelines, indications, and radiation dose risks. Int J Oral Maxillofac Implants. 2014;29(Suppl):55-77.
3. Misch CE. Contemporary Implant Dentistry. 4th ed. St. Louis: Elsevier; 2020.
4. Tyndall DA, Price JB, Tetradis S, et al. Position statement of the American Academy of Oral and Maxillofacial Radiology on selection criteria for the use of radiology in dental implantology. Oral Surg Oral Med Oral Pathol Oral Radiol. 2012;113(6):817-826.
5. 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(11):1243-1253.

6. Jacobs R, Salmon B, Codari M, Hassan B, Bornstein MM. Cone beam computed tomography in implant dentistry: recommendations for clinical use. BMC Oral Health. 2018;18(1):88.
7. The Academy of Prosthodontics. The glossary of prosthodontic terms: ninth edition. J Prosthet Dent. 2017;117(5S):e1-e105.
8. Dawood A, Patel S, Brown J. Cone beam CT in dental practice. Br Dent J. 2009;207(1):23-28.
9. Tischler M. In-office cone beam computerized tomography: technology review and clinical examples. Dent Today. 2008;27(6):102-106.
10. Zhang CN, Zhu Y, Fan LF, et al. Intra- and inter-observer agreements in detecting peri-implant bone defects between periapical radiography and cone beam computed tomography: a clinical study. J Dent Sci. 2021;16(3):948-956.
11. 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(5):443-464.
12. Arisan V, Karabuda ZC, Avsever H, Özdemir T. Conventional multi-slice computed tomography (CT) and cone-beam CT (CBCT) for computer-planned implant placement. Part I: relationship of radiographic gray density and implant stability. Clin Implant Dent Relat Res. 2013;15(6):893-906.
13. Parsa A, Ibrahim N, Hassan B, van der Stelt P, Wismeijer D. Bone quality evaluation at dental implant site using multislice CT, micro-CT, and cone beam CT. Clin Oral Implants Res. 2015;26(1):e1-e7.
14. Pauwels R, Jacobs R, Singer SR, Mupparapu M. CBCT-based bone quality assessment: are Hounsfield units applicable? DentomaxillofacRadiol. 2015;44(1):20140238.
15. Mah P, Reeves TE, McDavid WD. Deriving Hounsfield units using grey levels in cone beam computed tomography. DentomaxillofacRadiol. 2010;39(6):323-335.
16. Esposito M, Grusovin MG, Felice P, Karatzopoulos G, Worthington HV, Coulthard P. Interventions for replacing missing teeth: horizontal and vertical bone augmentation techniques for dental implant treatment. Cochrane Database Syst Rev. 2009;(4):CD003607.
17. Bornstein MM, Al-Nawas B, Kuchler U, Tahmaseb A. Consensus statements and recommended clinical procedures regarding contemporary surgical and radiographic techniques in implant dentistry. Int J Oral Maxillofac Implants. 2014;29(Suppl):78-82.
18. Pinheiro LR, Scarfe WC, Oliveira Sales MA, et al. 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(4):741-750.
19. Chang E, Lam E, Shah P, Azarpazhooh A. Cone-beam computed tomography for detecting vertical root fractures in endodontically treated teeth: a systematic review. J Endod. 2016;42(2):177-185.
20. Patel S, Dawood A, Ford TP, Whaites E. The potential applications of cone beam computed tomography in the management of endodontic problems. Int Endod J. 2007;40(10):818-830.
21. Fokas G, Vaughn VM, Scarfe WC, Bornstein MM. Accuracy of linear measurements on CBCT images related to presurgical implant treatment planning: a systematic review. Clin Oral

Implants Res. 2018;29(Suppl 16):393-415.

22. Ferrare N, Leite AF, Caracas HCPM, et al. Cone-beam computed tomography and microtomography for alveolar bone measurements. SurgRadiol Anat. 2013;35(6):495-502.
23. Gonzalez-Garcia R, Monje F. The reliability of cone-beam computed tomography to assess bone density at dental implant recipient sites: a histomorphometric analysis by micro-CT. Clin Oral Implants Res. 2013;24(8):871-879.
24. Javed F, Ahmed HB, Crespi R, Romanos GE. Role of primary stability for successful osseointegration of dental implants: factors of influence and evaluation. Interv Med Appl Sci. 2013;5(4):162-167.
25. Norton MR, Gamble C. Bone classification: an objective scale of bone density using the computerized tomography scan. Clin Oral Implants Res. 2001;12(1):79-84.
26. Ericsson I, Nilson H, Lindh T, Nilner K, Randow K. Immediate functional loading of Branemark single tooth implants. Clin Oral Implants Res. 2000;11(1):26-33.
27. Lundgren S, Cricchio G, Hallman M, Jungner M, Rasmusson L, Stenssson P. Sinus floor elevation procedures to enable implant placement and integration: techniques, biological aspects and clinical outcomes. Periodontol 2000. 2017;73(1):103-120.
28. Bergkvist G, Koh KJ, Sahlholm S, Klintström E, Lindh C. Bone density at implant sites and its relationship to assessment of bone quality and treatment outcome. Int J Oral Maxillofac Implants. 2010;25(2):321-328.
29. Akilimali A, et al. Artificial intelligence in dental radiology: a narrative review. Ann Med Surg. 2025;87(4):2212-2217.
30. Morgan N, Van Gerven A, Smolders A, de Faria Vasconcelos K, Willems H, Jacobs R. Convolutional neural network for automatic maxillary sinus segmentation on cone-beam computed tomographic images. Sci Rep. 2022;12(1):7523.