



# Visual grading experiments and optimization in CBCT dental implantology imaging: preliminary application of integrated visual grading regression

Ahmed Al-Humairi<sup>1</sup> · Ryan H. L. Ip<sup>2</sup> · Kelly Spuur<sup>3</sup> · Xiaoming Zheng<sup>3</sup> · Boyen Huang<sup>4</sup>

Received: 14 March 2021 / Accepted: 13 November 2021 / Published online: 5 January 2022  
© Crown 2022

## Abstract

This study uses a general formulation of integrated visual grading regression (IVGR) and applies it to cone beam computed tomography (CBCT) scan data related to anatomical landmarks for dental implantology. The aim was to assess and predict a minimum acceptable dose for diagnostic imaging and reporting. A skull phantom was imaged with a CBCT unit at various diagnostic exposures. Key anatomical landmarks within the images were independently reviewed by three trained observers. Each provided an overall image quality score. Statistical analysis was carried out to examine the acceptability of the images taken, using an IVGR analysis that was formulated as a three-stage protocol including defining an integrated score, development of an ordinal regression, and investigation of the possibility for dose reduction through estimated parameters. For a unit increase in the logarithm of radiation dose, the odds ratio that the integrated score for an image assessed by observers being rated in a higher category was 3.940 (95% confidence interval: 1.016–15.280). When assessed by the observers, the minimum dose required to achieve a 75% probability for an image to be classified as at least acceptable was 1346.91 mGy·cm<sup>2</sup> dose area product (DAP), a 31% reduction compared to the 1962 mGy·cm<sup>2</sup> DAP default dosage of the CBCT unit. The kappa values of the intra and inter-observer reliability indicated moderate agreements, while a discrepancy among observers was also identified because each, as expected, perceived visibility differently. The results of this work demonstrate the IVGR's predictive value of dose saving in the effort to reduce dose to patients while maintaining reportable diagnostic image quality.

**Keywords** Integrated visual grading regression · Dental implantology · Cone beam computed tomography · Medical imaging · Dose optimization · Image quality

## Abbreviations

ALQ	Anatomical landmark question	CT	Computed tomography
ANOVA	Analysis of variance	DAP	Dose area product
Az	Area under the receiver operating characteristic curve	DICOM	Digital imaging and communications in medicine
CBCT	Cone beam computed tomography	EAO	European Association of Osseointegration
CI	Confidence interval	Eq.	Equation
		exp	Exponential function
		FOV	Field of view
		IIQ	Integrated image quality
		IVGR	Integrated visual grading regression
		kVp	Peak kilovoltage
		LF	Lingual foramen
		ln	Natural logarithm
		mA	Milliamperere
		MC	Mandibular canal
		MF	Mental foramen
		mGy·cm <sup>2</sup>	Milligray times square centimeter
		MRI	Magnetic resonance imaging
		OIQ	Overall image quality

✉ Ahmed Al-Humairi  
a.alhumairi@uq.edu.au

<sup>1</sup> School of Dentistry, The University of Queensland, Herston, QLD, Australia  
<sup>2</sup> School of Computing, Mathematics and Engineering, Charles Sturt University, Wagga Wagga, NSW, Australia  
<sup>3</sup> School of Dentistry and Medical Sciences, Charles Sturt University, Wagga Wagga, NSW, Australia  
<sup>4</sup> Department of Primary Dental Care, University of Minnesota School of Dentistry, Minneapolis, MN, USA

ROC	Receiver operating characteristic
RT	Right-side
sd	Standard deviation
SE	Standard error
VGA	Visual grading analysis
VGAS	Visual grading analysis score
VGC	Visual grading characteristic
VGR	Visual grading regression

## Introduction

Cone beam computed tomography (CBCT) is used for the evaluation and diagnosis of disease with routine use continuing to increase (Anderson et al. 2014). However, the luxury of using this diagnostic tool comes together with a radiation dose risk to the patient. Although low dose protocols have been developed to address this matter, there is a lack of consistency among those protocols and clinical uses in current CBCT models (Nemtoi et al. 2013; Anderson et al. 2014), particularly for dose risk and image quality (Carter et al. 2008).

Visual grading experiments have been frequently used to assess image quality. In these experiments, each image is assessed by multiple observers for optimal diagnostic performance and is assigned a score reflecting the extent of image quality. An example is to evaluate the subjective image quality with visual grading analysis (VGA) (Hidalgo Rivas et al. 2015; Kadesjö et al. 2015; Liang et al. 2010; Pittayapat et al. 2013; Vandenberghe et al. 2007). To optimize the radiation dose used in the clinical setting, it is important to relate the physical image evaluation to the subjective image quality. Often, the scores from VGA are defined on a 3-, 5-, or 7-step Likert-type scale which is a widely used psychometric scale (Likert 1932). For example, when a 5-step scale is used, it could be presented as 1 for “Clearly not visible” to 5 for “Clearly visible”. In this sense, the scores are defined in an ordinal scale, meaning that they have a natural ordering but the differences between 1 and 2 may not be the same as those between 2 and 3. The ordinal nature poses a challenge to researchers as it requires some special techniques to handle.

To analyze data from visual grading experiments, the method of visual grading analysis score (VGAS) is often used—it simply calculates the average score across all criteria and observers (Kadesjö et al. 2015; Månsson 2000). The scores are then plotted against the explanatory variables or compared between different groups using statistical methods such as *t*-test and analysis of variance (ANOVA). Due to the ordinal nature of the data, any methods involving calculation of the means are inappropriate, as this would assume that the data are interval or ratio in nature.

In making the image analysis statistically valid, a visual grading characteristic (VGC) method was developed and

formulated based on the receiver operating characteristic (ROC) method (Båth and Månsson 2007). However, VGC is limited to comparing two parameters at a time (Zarb et al. 2015; Zheng et al. 2016). When assessing the effects of more than two parameters, researchers can use the visual grading regression (VGR) method to handle ordinal data (Zarb et al. 2015; Zheng et al. 2016; Smedby and Fredrikson 2010; Smedby et al. 2013; Saffari et al. 2015; Agresti 2010). With this approach, the probability of the response variable being less than a certain score is modeled and the simultaneous effects (fixed and/or random) of several explanatory variables can be assessed (Hedeker and Gibbons 1994). Further, it can be easily performed with almost all modern statistical software.

When observers are asked to give only one score for each image, VGR can be applied directly, and the differences among observers can be captured by incorporating random effects (Zarb et al. 2015). In some visual grading experiments, observers are asked to give more than one score for each image, such as to assess the visibility of several (*n*) anatomical landmarks and an overall quality of an image. The scale of the scores for the visibility (e.g., a 5-step scale) of individual anatomical landmarks may be different from that for the overall image quality (e.g., a 3-step response representing the acceptability of the image for clinical use). In the above example, there are a total of *n* + 1 response variables (*n* = anatomical landmarks plus an overall image quality response) for each image, and they are statistically dependent. If a low score is given to an anatomical landmark, other anatomical landmarks will be likely to receive low scores as well. In this regard, the data structure is multivariate ordinal. The ordinal regression model, typically assuming independence among response variables, may become questionable if used in the scenario. Although a multivariate ordinal regression model is more appropriate to handle this kind of data (Liu and Hedeker 2006), practitioners may find it hard to understand and interpret the results.

In multivariate statistics, to handle the challenges brought by the dependency structure among the variables, it is common to reduce the dimension of the multivariate data. In terms of visual grading experiments, one can reduce the dimension of the multivariate ordinal data by defining an integrated image quality (IIQ) score for each image based on all scores given, which technically creates a new response variable (Hidalgo Rivas et al. 2015; Al-Humairi et al. 2016a, b). However, in previous IIQ applications, the effects of explanatory variables have not been quantified and inter-observer variabilities have not been considered (Hidalgo Rivas et al. 2015; Al-Humairi et al. 2016a, b). In this paper, a new method is proposed which combines the IIQ and VGR methods to manage the multivariate ordinal data arising from visual grading experiments. This method is named here as the integrated visual grading regression (IVGR) model.

Planning for dental implant surgery was used to investigate this new method. Dental implant surgery is often considered as an elective procedure; and updated radiographs and/or CBCT images are listed as surgical safety requirements of this procedure (Bidra 2017). While CBCT is useful “to evaluate the position of the implant and its surrounding structures, and to determine whether the implant is removed, following dental implant surgery” (Kim et al. 2020), the effective dose of dental CBCT units differs markedly (Rios et al. 2017). As recommended by the European Association of Osseointegration (EAO), the imaging technique chosen should be optimized to provide the relevant diagnostic information with the least radiation dose (Harris et al. 2012). To the best of the authors’ knowledge, there are no consensus guidelines of dose or image quality for dental implantology. Hence, the aims of this study were to (1) provide a general formulation and application of the IVGR method and (2) assess and predict possible radiation dose reduction, using a set of CBCT scan data related to anatomical landmarks for dental implantology.

## Materials and methods

### Application

#### Experimental setup

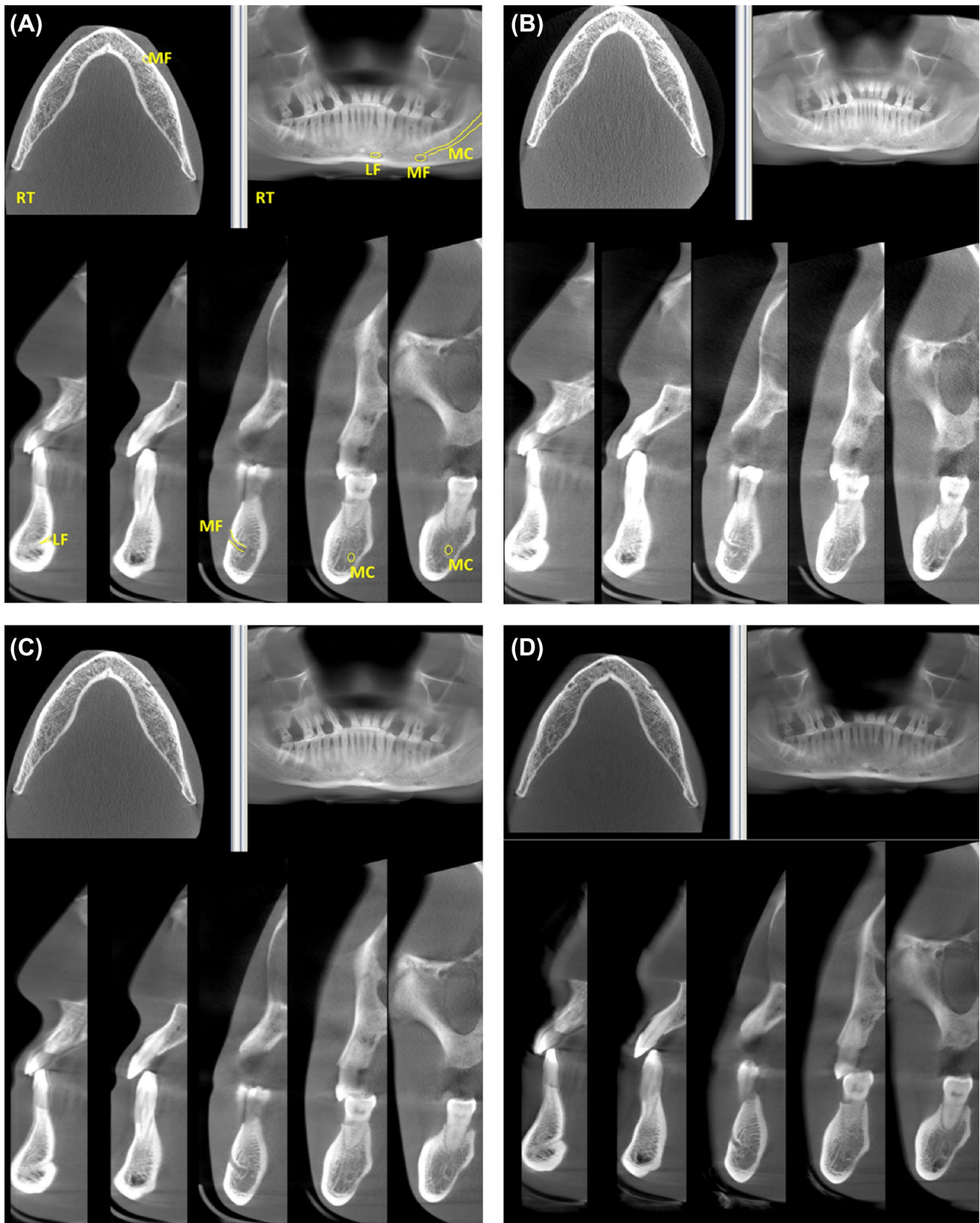
A skull phantom comprising a dry adult human skull embedded in Plexiglas-simulating soft tissue (3 M MRI CT Phantom Real Human Skull), was used. The skull was imaged with a CBCT unit (Planmeca Promax 3D Max) operated at 70 kVp and 8, 10 and 12 mA, 80 kVp and 4, 6, 8, 10, and 12 mA, and 96 kVp and 4, 6, 8, 10 and 12 mA, using a large field of view (FOV) to the maxillofacial area (Fig. 1), with 32 s scan time and a voxel size of 200  $\mu\text{m}$ . The exposure parameters were pre-determined by the manufacturer; no manual adjustment was performed. The radiation dose was recorded as a dose area product (DAP), in  $\text{mGy}\cdot\text{cm}^2$ , which was extracted from the data embedded in each image. Ethics and radiation safety approval was granted by the Charles Sturt University Human Research Ethics and Radiation Safety Committees (Reference Number: 414/2013/01).

Volume datasets were stored in the Digital Imaging and Communications in Medicine (DICOM) format. Axial, parasagittal, and three-dimensional reconstructed images through the area of interest and a cross-section of the posterior mandible through the middle of the prospective implant position were selected for review. The posterior mandibular region was selected for investigation because the anatomy of this site presents a higher risk for dental implant surgery and therefore is of diagnostic importance (Froum et al. 2011). All images were independently reviewed in blinded random

order by three separate observers who were experienced dental clinicians capable of providing surgical service for dental implantology and trained to interpret dental CBCT images for the study task. After each observer completed a calibration training session, they evaluated the visibility of anatomical landmark quality and overall image quality as per routine pre-surgical assessment of implant placement. The observers ranked seven anatomical landmark questions (ALQ, Table 1) on a five-point rating scale; 1: definitely it is not clearly visible, 2: probably it is not clearly visible, 3: indecisive whether that is visible or not, 4: probably it is clearly visible, and 5: definitely it is clearly visible. In addition, an overall image quality (OIQ) score for the pre-surgical implant placement image on a three-point rating scale (poor, acceptable and clear) was recorded. A VGA method was used because it was believed that the decision on clinical adequacy or unacceptability of image quality for clinical purposes should authentically rest with the observers. The observers were allowed to record their subjective opinion regarding the visibility of the anatomical landmarks or structures relevant to the clinical indication. The absolute VGA (a score given to an image by the observer based on their experience), was adopted in the study. The absolute VGA is a preferred method for quantification of subjective opinions, since Zarb et al. (2015) have reported that the images were comparable to each other when using the absolute VGA for CT scan optimization on head scanning.

Image analysis was based on the requirement of the image at the pre-surgical stage of implant placement in the mandible; for the purpose of this study, only the left side of the mandible was assessed. In each view up to four slices were presented so that all ALQs were able to be adequately visualized and assessed. The scores given by all observers were documented for each exposure. Further, the observers made an overall analysis of the images indicating whether the images were acceptable for the diagnostic task on the site of the pre-surgical implant placement.

Images were evaluated under routine clinical viewing conditions in the reporting setting. Observers were instructed to rest their eyes if they felt they were suffering eye strain. In addition, observers were required to replicate their clinical work environment and wear glasses if used as well as changing the display window width and level or using any magnification methods if they typically used them. They were also allowed to adjust the brightness and contrast as they thought appropriate. Each observer was required to assess thirteen acquisitions along with five duplicated and randomly presented slices for testing intra-observer reliability. The inter-observer reliability was measured by comparing the scores between observers. The kappa statistic was used to test the intra- and inter-observer agreement in this study.



**Fig. 1** Examples of cone beam computed tomography (CBCT) images and the anatomical landmarks used in this study: **(A)** CBCT image at 80 kVp and 10 mA, showing the maxillofacial landmarks

**(MC: mandibular canal, MF: mental foramen, LF: lingual foramen, RT: right-side), (B)** CBCT image at 70 kVp and 12 mA, **(C)** CBCT image at 80 kVp and 10 mA, **(D)** CBCT image at 96 kVp and 10 mA

**Table 1** Anatomical landmark questions

Anatomical landmark questions	Evaluation criteria <sup>a</sup>
Bone width	Image quality sufficient for an adequate depiction of bone width in different views
Bone height	Image quality sufficient for an adequate depiction of bone height for implant insertion with reference to the proximity of the relative anatomical structure
Mandibular canal, molar area	Clearly visible outline of the mandibular canal in the molar area
Mental foramen	Clearly visible mental foramen
Lingual foramen	Clearly visible lingual foramen
Cortical bone	Image quality sufficient for an adequate depiction of the cortical bone
Trabecular bone	Image quality sufficient for an adequate depiction of the trabecular bone

<sup>a</sup>Established based on Goto et al. (2007), Koizumi et al. (2010), and Hofmann et al. (2014)

## Integrated visual grading regression (IVGR)

An IVGR analysis can be formulized as a three-stage protocol. In Stage 1, an integrated score based on certain criteria must be defined. Each image is then assigned an integrated score to represent its overall image quality. In Stage 2, an ordinal regression model is fitted using the integrated score in Stage 1 as the response variable on the explanatory variables in the study. To capture the variabilities of the observers, it is suggested to include the observers as a random effect. In Stage 3, the effects of the explanatory variables are assessed based on the fitted model. This stage varies depending on the field of application and the aim of the experiment. To use IVGR analysis, the integrated scores must be ordinal in nature. If the integrated scores belong to interval or ratio data, VGAS could be used instead. The idea of IVGR is formulized in generic terms below.

Consider an experiment involving  $K$  observers, each was asked to give  $q$  scores for each of  $J$  images. Denote by  $Y_{ijk}$  the  $i$ th score for the  $j$ th image given by observer  $k$  for  $i = 1, \dots, q; j = 1, \dots, J; k = 1, \dots, K$ . The scales of  $Y_i$  for different  $i$  may differ. The quality of the image is hypothesized to be affected by  $p$  explanatory variables  $X_1, X_2, \dots, X_p$ .

### Stage 1: Defining an integrated score

In this stage an integrated score  $Z_{jk}$  is needed for image  $j$  assessed by observer  $k$  based on the  $q$  scores given. In general,  $Z_{jk}$  is a function of  $Y_{ijk}$  for all  $i = 1, \dots, q$

$$Z_{jk} = f(Y_{1jk}, Y_{2jk}, \dots, Y_{qjk}) \quad (1)$$

where  $f$  is a function used to classify the images into  $T$  ordinal categories.

For simplicity, assume the categories are labeled by  $1, 2, \dots, T$ . The new variable  $Z$  is univariate ordinal in nature. Thus, the dimension of the data can be reduced from  $q$  to one, bypassing the difficulties of handling multivariate ordinal data.

### Stage 2: Ordinal regression

Here an outline of ordinal regression is provided. Readers are referred to statistical texts such as Agresti (2007), Powers and Xie (2008) and Kleinbaum et al. (2007) for more details.

As  $Z_{jk}$  now is univariate ordinal in nature, usual ordinal regression approaches can be used. Common choices of the link functions include logit and probit (Agresti 2007). In the present work focus is on the logit link below as it is often found to be more intuitive and easier to interpret (Dow and Endersby 2004). In particular, one models the natural logarithm of the odds of obtaining  $Z_{jk}$  not greater than a particular level  $c$  against  $Z_{jk}$  greater than  $c$  using a regression equation for  $c = 1, 2, \dots, T - 1$ , assuming there are  $T$  levels. Putting the  $p$  explanatory variables as fixed effects and the observer random effects into the model, the ordinal regression model takes the form

$$\begin{aligned} \text{logit}[P(Z_{jk} \leq c)] &= \ln \left[ \frac{P(Z_{jk} \leq c)}{P(Z_{jk} > c)} \right] \\ &= \alpha_c - \sum_{m=1}^p \beta_m X_{mj} - \delta_k, c = 1, 2, \dots, T - 1 \end{aligned} \quad (2)$$

where  $\alpha_c$  is the threshold parameter,  $\beta_m$  is the coefficient for  $X_m$ , often called the effect parameter, and  $\delta_k$  is the random effect for observer  $k$ .

If one defines  $Z_{jk}$  on a binary scale, the ordinal regression model reduces to a logistic regression model. In Eq. (2), all  $X_m$  are assumed to be continuous. If some of them are qualitative factors, indicator variables can be used accordingly. The `clmm2` command from the ordinal package of R (R Core Team 2020) is capable of fitting the above model (Christensen 2019, 2015).

### Stage 3: Model interpretation

With the fitted parameters, given the values of the explanatory variables, one would calculate the odds ratio or the probability that  $Z_{jk}$  is classified into a particular category. If  $\beta_m > 0$ ,  $Z$  tends to be higher at higher levels of  $X_m$ , when all other explanatory variables remain unchanged. In particular,

$\exp(\beta_m)$  represents the odds ratio of  $Z$  being rated at a higher category when  $X_m$  increases by one unit while all other explanatory variables remain unchanged. The threshold parameters  $\alpha_c$  represent the log-odds of  $Z_{jk}$  being classified into category  $c$  or below when the image is assessed by an ‘average’ observer (so that  $\delta_k = 0$ ), and all explanatory variables  $X_m$  equal to zero. In practice, these threshold parameters may have no meaningful interpretation when it may not be sensible to have all  $X_m$  equal to zero. As described in Agresti (2007), the  $\beta'$ ’s are usually the parameters of interest. In terms of probabilities, from Eq. (2), one would calculate the probability that  $Z_j$  is greater than  $c$  for any  $c$  between 1 and  $T - 1$  (inclusive) when an image  $j$  is assessed by an ‘average’ observer (so that  $\delta_k = 0$ ) as

$$P(Z_j > c) = \frac{1}{1 + \exp[\hat{\alpha}_c - \sum_{m=1}^p \hat{\beta}_m X_{mj}]}, \quad (3)$$

where  $\hat{\alpha}$  and  $\hat{\beta}$  denote the estimates of the corresponding parameters.

As another point of view, assume one wishes to minimize  $X_1$  while maintaining a probability  $p_0$  (where  $0 < p_0 < 1$ ) of the integrated score being greater than  $c_0$ , the minimum of  $X_1$  can be found by

$$X_1^{\min} = \frac{\hat{\alpha}_{c_0}}{\hat{\beta}_1} - \frac{\sum_{m=2}^p \hat{\beta}_m X_{mj}}{\hat{\beta}_1} - \frac{1}{\hat{\beta}_1} \ln \left( \frac{1}{p_0} - 1 \right). \quad (4)$$

Equation (4) was used to find the minimum radiation dose level in the present application.

## Data analysis

As above, denoted by  $Y_{ijk}$  the score of the  $i$ th question of the  $j$ th image assessed by observer  $k$ . Here,  $Y_1$  to  $Y_7$  represented the scores for the seven ALQs, and  $Y_8$  the OIQ score. Out of the eight scores, the last one,  $Y_8$ , is perhaps the most important one and has to be treated differently. It is natural to assume that an image should be at least acceptable for clinical use. Therefore,  $Y_8 \geq 2$  is required. The integrated image quality for image  $j$  assessed by observer  $k$  on a 4-step scale was defined as follows:

$$Z_{jk} = \max \left\{ 0, \sum_{i=1}^7 1 \{ Y_{ijk} \geq 4 \} - 4 \right\} \times 1 \{ Y_{8jk} \geq 2 \} \quad (5)$$

where  $1\{A\}$  is the indicator function which takes a value of 1 if the condition  $A$  is satisfied; and a value of 0 otherwise.

It can be easily verified that  $Z_{jk}$  takes a value from the set  $\{0, 1, 2, 3\}$ . If the image is not acceptable for clinical use ( $Y_{8jk} < 2$ ) and/or less than five of the seven anatomical landmarks scored ‘‘4’’ or above, then  $Z_{jk} = 0$ , representing a poor integrated image quality. If the image is acceptable for clinical use ( $Y_{8jk} \geq 2$ ),  $Z_{jk}$  will depend on the number of

anatomical landmarks scored ‘‘4’’ or above. Naturally, the more the anatomical landmarks scored ‘‘4’’ or above, the better the image quality reflected by  $Z_{jk}$ . Overall speaking, one could interpret the image quality as ‘‘poor’’ if  $Z_{jk} = 0$ ; ‘‘acceptable’’ if  $Z_{jk} = 1$ ; ‘‘good’’ if  $Z_{jk} = 2$ ; and ‘‘excellent’’ if  $Z_{jk} = 3$ .

In general, the concept of integrated score is flexible in the sense that Eq. (5) can be modified easily to cater different needs in various applications. As given in Eq. (1), any sensible choice of function  $f$  could be used. It is possible to include more or fewer ranks, as well as making the criteria more or less stringent. However, caution must be taken especially if one wishes to make the criteria less stringent. In medical studies, it is suggested to define rules which are tighter rather than looser.

Without having other explanatory variables, the level of dose is the sole explanatory variable, denoted by  $X$ . Natural logarithm transformation was applied on  $X$ , as suggested by Smedby et al. (2013). With observers considered as the random effects, the ordinal regression model admits the form

$$\text{logit}[P(Z_{jk} \leq c)] = \alpha_c - \beta \ln(X_j) - \delta_k, \quad c = 0, 1, 2. \quad (6)$$

It is demanded to have a probability of at least 75% for an image being classified as at least acceptable (Jones et al. 2015; Favazza et al. 2015; Schaefferkoetter et al. 2015; Prasad et al. 2002), namely  $P(Z_{jk} > 0) = P(Z_{jk} \geq 1) \geq 0.75$ , when the image is assessed by an ‘average’ observer. In other words, following Eq. (4), the minimum level of dose required can be calculated as

$$X_{\min} = \exp \left[ \frac{\hat{\alpha}_0}{\hat{\beta}} - \frac{1}{\hat{\beta}} \ln \left( \frac{1}{0.75} - 1 \right) \right] \quad (7)$$

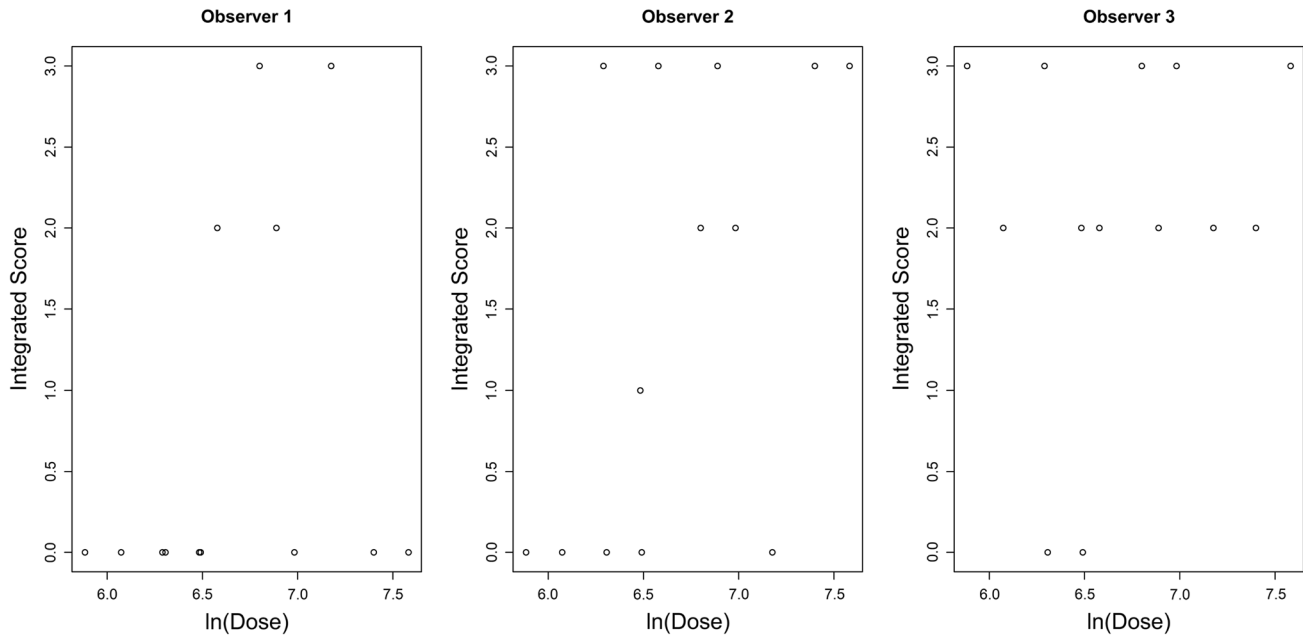
If another probability of detection threshold (e.g., 50%) is used, one can replace the number of 0.75 with the corresponding value in Eq. (7).

## Results

As described earlier, the main purpose of the paper is to apply IVGR to CBCT and investigate the potential for dose reduction while maintaining acceptable image quality. Following this, the exposure parameters and observers’ scores are presented in Table 2, and the integrated scores against the logarithm of dose level according to the observers in Fig. 2. In general, a higher dose level led to a higher integrated score even though low scores were occasionally given by Observer 1 for high dose levels and high scores were sometimes given by Observer 3 for low dose settings. The kappa value of the intra-observer reliability for each observer is displayed in Table 3. These values indicate moderate to

**Table 2** Exposure parameters and observer scores (N: the number of anatomical landmarks being scored as 4—probably it is clearly visible, or 5—definitely it is clearly visible; OIQ: the overall image quality score based on the three-point rating scale)

Exposure Parameters	kVp mA	70			80				96					
		8	10	12	4	6	8	10	12	4	6	8	10	12
Observer 1	N	2	4	2	2	4	6	7	3	5	6	7	2	4
	OIQ	1	1	1	1	1	2	2	1	1	2	2	1	1
Observer 2	N	1	2	1	5	7	7	6	6	5	7	5	7	7
	OIQ	1	1	1	1	2	2	2	2	2	2	1	3	2
Observer 3	N	6	5	5	7	7	6	7	7	6	6	6	6	7
	OIQ	2	1	1	2	2	3	3	3	2	2	2	2	2



**Fig. 2** Plot of integrated scores against logarithm of dose level by observers

**Table 3** Kappa values demonstrating the intra-observer (diagonal) and inter-observer (off-diagonal) agreements

Observer	1	2	3
1	0.477	0.322	0.261
2		0.436	0.468
3			0.777

substantial agreements (Landis and Koch, 1977). The kappa values of the inter-observer reliability ranged from 0.261 to 0.468 (Table 3). These values indicate fair to moderate agreements (Landis and Koch 1977). As visual grading is a subjective task, it is natural to see a lower inter-observer reliability (Lee et al. 2019), compared to intra-observer reliability. The between-observer variabilities were captured as random effects in the model.

Table 4 shows the estimated parameters for the ordinal regression model described by Eq. (6). From the p-values, both  $\alpha_2$  and  $\beta$  are significant at 5%. Since  $\ln(X)$  cannot be 0, the estimated threshold parameters have no meaningful

**Table 4** Result of ordinal regression analysis

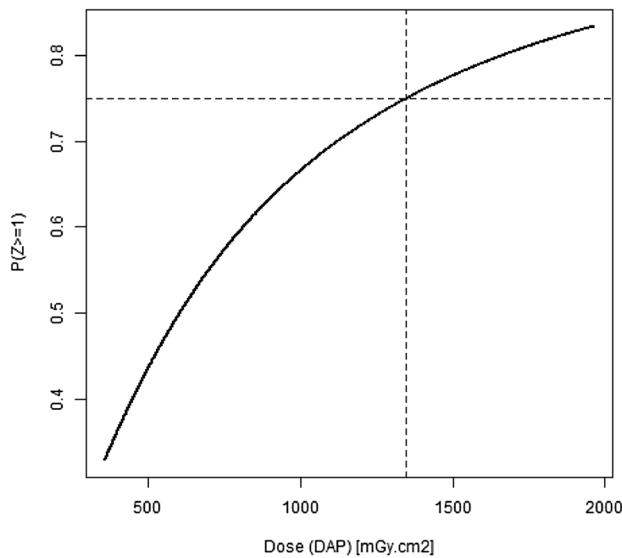
Parameter <sup>a</sup>	Estimate	SE	p value	95% Wald confidence interval <sup>b</sup>
$\alpha_0$	8.781	4.642	0.059	(− 0.301,17.863)
$\alpha_1$	8.909	4.649	0.055	(− 0.185,18.004)
$\alpha_2$	10.154	4.717	0.031*	(0.934,19.374)
$\beta$	1.371	0.693	0.048*	(0.016, 2.727)
$sd(\delta_k)$	0.568			

<sup>a</sup> $\alpha$  and  $\beta$  represent the threshold and effect parameters, respectively, in Eq. (2), and  $sd(\delta_k)$  represents the standard deviation of the random effects

<sup>b</sup>A 95% Wald confidence interval for a parameter is calculated as estimate  $\pm 1.96 \times$  standard error (SE) (Agresti 2010)

\*  $p < 0.05$

interpretation. However, these were used to derive the probability of an image having a particular integrated score. The coefficient of  $\ln(X)$  is positive indicating that a higher level



**Fig. 3** Plot of probability of  $Z_{jk}$  being greater than or equal to 1 against the dose level; DAP dose area product

of dose increased the image quality, *id est*, the image is more likely to be classified in higher categories. Specifically, when  $\ln(X)$  is increased by 1 (that is, when the level of dose is multiplied by  $\exp(1) = 2.718$ ), the odds ratio of  $Z_j$  being rated in a higher category is  $\exp(1.371) = 3.940$  (95% confidence interval (CI): 1.016–15.280).

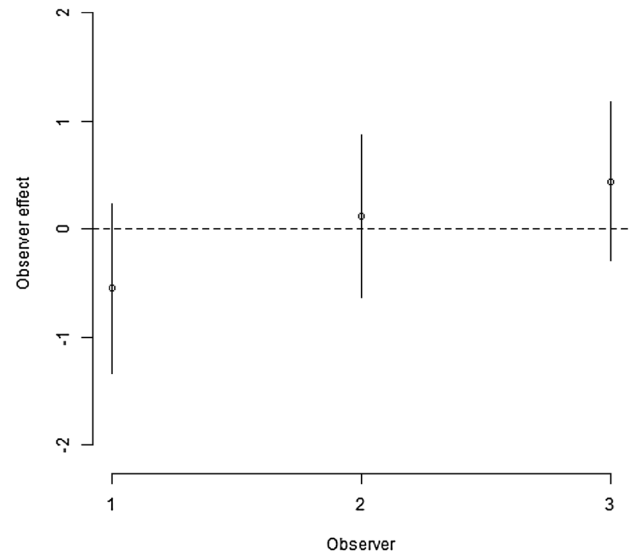
Figure 3 shows the plot of  $P(Z_{jk} \geq 1)$  against the dose level. The minimum dose level required to achieve a probability of 75% for an image being classified as at least acceptable, when assessed by an ‘average’ observer (that is, when  $\delta_k = 0$ ) is 1,346.91 mGy·cm<sup>2</sup> DAP, a 31% reduction compared to 1,962 mGy·cm<sup>2</sup> DAP, which is the default dosage of the CBCT unit used (Al-Humairi et al. 2016a).

The estimated random effects can also be extracted from the model. Figure 4 shows the estimated modes and the 95% CIs for each of the observers. Among the observers, Observer 1 tends to give the lowest rating while Observer 3 tends to give the highest rating. Again, such a discrepancy is not unexpected as each person perceives visibility differently.

## Discussion

### Integrated visual grading regression (IVGR)

This study has reported for the first time the use of a statistically feasible IVGR method to analyze the multivariate ordinal data of subjective image quality in the context of CBCT clinical pre-surgical planning for dental implant placement. The relevance of human perception and cognition was highlighted by this work. Even though some researchers have



**Fig. 4** Observer effects given by 95% confidence intervals based on the estimated variances of random effects

assessed subjective image quality in CBCT dental implantology imaging, none of them has applied IVGR to manage the ordinal data obtained from the observers’ grading scores (Lofthag-Hansen et al. 2011; Dawood et al. 2012; Alawaji et al. 2018; Park et al. 2019; Shelley et al. 2011). Previous studies have had different focuses such as the effects of exposure parameters (Lofthag-Hansen et al. 2011; Dawood et al. 2012; Alawaji et al. 2018; Park et al. 2019), FOVs (Lofthag-Hansen et al. 2011), and imaging systems (Shelley et al. 2011) on subjective image quality. Although higher inter-observer agreements have been reported by Shelley et al (2011) and Park et al (2019), their papers and Lofthag-Hansen et al (2011) also presented inconsistent intra-observer agreements among the observers, similar to the present data. This study also suggested a 31% reduction of the CBCT radiation dose from the manufacturer’s recommendation and this agreed with Alawaji et al (2018) who considered the possibility of reducing dose by 30%. Dawood et al (2012) estimated a dose reduction up to 87.5% from the default setting even though the utilization of three-dimensional reconstruction would be compromised. While the authors of the present paper acknowledge the contribution of the earlier papers on optimization of radiation dose in dental implantology imaging, the present study has added the value of the IVGR method to this field.

In the present model, the dose value of 1346.91 mGy·cm<sup>2</sup> DAP was predicted under the assumption of 75% probability for an image being classified at least acceptable for diagnosis. In clinical ROC studies, a value of  $A_z = 0.75$  is generally accepted as a common value and anything above ( $A_z > 0.75$ ) is considered as superior (Metz 1989). In psychophysics studies, a 50% probability of detection is generally



considered as the threshold (Krantz 2012). The predicted dose value would be much lower than the DAP value of 1346.91 if the threshold probability was set at 50% in the present model. The predicted dose value is therefore not a threshold dose but an indicative dose value that is acceptable for clinical practice. Acceptable image quality, that supports clear identification of anatomical structures as well as the morphology, dimension and quality of the bone, is required for development of an acceptable image quality protocol. Development of such a protocol has further potential to promote dose reduction (SEDEXCT 2012). This is a reason why the image quality of selected anatomical landmarks as well as trabecular and cortical bone were evaluated in the present study. An earlier paper has also reported the influence of optimization protocols on the associated image quality of cortical and trabecular patterns (Koizumi et al. 2010). In general, image quality is assessed with established criteria for the visibility of key landmarks (Attard et al. 2018). While objective methods are repeatable, defining the image quality is clinical task based (Barrett et al. 2015). Owing to a gap between subjective and objective assessment methods for image quality analysis (Hidalgo-Rivas et al. 2014), visibility of anatomical landmarks alone cannot be considered as an adequate performance indicator (Zanca et al. 2012). The present study aimed to explore potential for implementation of an IVGR assessment method on maxillofacial CBCT images but not to suggest this as a superior substitute for objective assessment methods.

### Statistical considerations

The key to the success of visual grading experiments is defining an integrated score. In general, clinical image quality is criteria based, and there is considerable known inter- and intra-observer variability in VGA for specific criteria of a specific image, which is the main challenge in employing VGA in quantifying clinical image quality. Undoubtedly, ordinal regression should be employed in VGA for ordinal data. It is common to incorporate random effects into the model to capture the heterogeneity between the observers, especially when large variability exists (Drikvandi and Noorian 2019).

Smedby and Fredrikson (2010) stated that it is not statistically acceptable to use a common statistical method relying on least-squares estimates such as *t*-test and ANOVA on ordinal type data. VGA is an easy and straightforward approach, but the statistical analysis of the scoring data has some limitations and there is a lack of consistency in the choice of the methods. The scoring data are usually non-linear numerical values and consequently they do not fit the parametric statistical methods such as ANOVA. To address the issue, the methods of non-parametric visual grading

(Båth and Månsson 2007) and VGR (Zarb et al. 2015; Zheng et al. 2016; Smedby and Fredrikson 2010; Smedby et al. 2013; Saffari et al. 2015) provide easier and more practical applications which have been proposed by previous researchers. VGR methodology is in agreement with the present study developed using methodology which introduces the concept of IIQ by adding a clinical question (an IOQ that evaluated whether the overall image quality was acceptable for a pre-surgical assessment of a dental implant) to it. VGR methodology can be used for multivariate ordinal regression, such as various anatomical landmarks and overall scores in this study. The difficulty of using general VGR in the present study is that the individual landmark scores and overall scores are not generally independent from each other. The IIQ was thus to select the critical components of the multivariate scores to form a single integrated score for the VGR.

### Image quality

Linear, logarithmic, or logistic (non-linear) functions have all been reported in the literature for the relationship between diagnostic image quality and dose. The logistic function can be considered as the united function for all of them (Zheng 2017). Psychometric factors affect the evaluation of image quality because image quality involves the interaction between human attitude and image detail. The observer will score the image attributes in relation to their agreement about whether they are clearly visualized. This is called the self-efficacy theory, which was reinforced and linked to image quality as described by Mraity et al (2014). The present approach of validation of the psychometric scale of image quality and developing an image quality method that answers the principal clinical questions agrees with this theory. A psychometric approach explains the other application of the proposed method to eliminate the disagreement between observers and link it to the psychometric approach while answering the clinical questions. Apart from dose levels, the proposed approach is flexible in the sense that other psychometric factors such as decision thresholds can be included in the model as extra explanatory variables, provided that they were recorded during the experiments.

The optimization methods in dental radiology focus on providing an image that fits the clinical purpose adequately while minimizing radiation exposure to the patient. An image quality index was defined using a descriptor in an ordinal scale based on subjective evaluation of the visual data contained within the image. Therefore, it is widely agreed that the term *adequate* or *acceptable image quality* indicates a satisfactory answer to the primary clinical question (Månsson 2000; Båth and Månsson 2007).

## Limitations and future directions

A limitation of this study is the small number of observers used. Although it would be ideal to determine the minimum number of observers required for dose optimization, in the statistics literature, practical methods for estimating the minimum sample size for general ordinal regression problems with random effects are yet to be established. Based on simulations, Ali et al. (2016) recommended using a minimum of 50 groups (observers in the present application) to achieve a power of 80% or above; and Bauer and Sterba (2011) suggested that the maximum likelihood estimators were least biased for data with at least 100 groups. However, the recruitment of such a large number of observers in visual grading experiments is not practical. Moreover, it was challenging to recruit fully registered specialists to participate in the current study due to the significant time required for observation and scoring. As a recent study has used only three observers (Almashraqi et al. 2017), with adequate pre-training provided to the observers, the small number is considered acceptable for this research purpose. Intra- and inter-observer agreements reported in this study were also consistent with those of previous studies (Hidalgo Rivas et al. 2015; Heetveld et al. 2005). Although the present observers were experienced clinicians specializing in dental implantology and completed a calibration training session prior to participation, more extensive trainings may improve the agreements.

Further, the conduction of this experiment on a single device only, the use of a single phantom only and the large FOV, are also limitations of this study. The use of a large FOV to the maxillofacial area replicates some clinical scenarios where an evaluation of several edentulous areas for implant placement is indicated, in preference to multiple radiation exposures. As smaller FOVs are more often used in dental implantology, these should also be considered in future studies.

The quality of a radiographic image is essentially determined by the observers' opinions (Sund et al. 2004) which are based on their individual experience and other technical parameters. A previous study emphasized observer perception and cognition as a relevant factor in image quality assessment (Kundel 2015). Image perception is considered as a unified realization of the contents of the image signal (displayed image) and cognition is the ability to explain the connotation of the displayed images in the context of medical scenarios. The psychological (human visualization and perception) and physical (anatomical landmark) elements combine to inform the evaluation of the image (Rossmann and Wiley 1970). As making a clinical judgment is a complex decision-making process, superior resolution of anatomical and physical landmarks on the displayed image can influence observer variability by focusing the observer on

certain structures within the image (Thornbury et al. 1978). As shown in Fig. 2, a high dose level may not necessarily result in higher integrated score. On one hand, this may be due to the subjective nature of the evaluation tasks—observers' own personal likings. On the other hand, it may indicate that the dose level is not the sole explanator. While the effect of dose was found to be statistically significant, the small number of observers led to a relatively large standard error, causing wide confidence intervals for the effect and threshold parameters. Thus, to validate the dose level proposed in this study, a larger scale investigation using more observers and additional explanatory variables such as extra anatomical landmarks and pathologies is indicated.

As medical imaging is an essential tool used in the diagnosis and treatment planning of various health conditions (Sakata-Goto et al. 2012; Spuur 2019; Tanny et al. 2018), IVGR may be useful in creating optimization protocols to further benefit the safety of patients by establishing minimum acceptable dose levels for diagnostic imaging and reporting in other medical imaging modalities. Future investigations including the fields of orthopedics, mammography, traumatology and orthodontics are indicated.

## Conclusion

This study has reported a preliminary and achievable application of IVGR in CBCT dental implantology imaging. Within the limitations of this study, the authors have highlighted the conundrum of the putative statistical analysis of visual grading scoring. Therefore, part of the conclusion of this study clarifies the feasibility of the derived IVGR method. With a 31% dose reduction estimated, this study has also demonstrated that IVGR can be a valuable method for dose minimization, which may be used in the future to predict optimization methods for specific clinical tasks and develop low dose protocols. This conclusion is pertinent for clinicians and researchers, as it highlights the need to underpin research methodology with carefully controlled experiments for the potential reduction of radiation dose. Further investigations with more human observers are indicated to validate the IVGR model in other clinical applications including conventional CT and planar radiographic imaging.

## Declarations

**Conflict of interest** The authors have no relevant conflicts of interest to disclose.

## References

- Agresti A (2007) An introduction to categorical data analysis. Wiley, New York
- Agresti A (2010) Analysis of ordinal categorical data. Wiley, Hoboken
- Alawaji Y, MacDonald DS, Giannelis G, Ford NL (2018) Optimization of cone beam computed tomography image quality in implant dentistry. *Clin Exp Dent Res* 4:268–278. <https://doi.org/10.1002/cre2.141>
- Al-Humairi A, Zheng X, Ip HL, El Masoud B (2016) Radiation dose image quality optimization in dental implantology. In: International conference on applied mathematics, simulation and modelling. Atlantis Press, Beijing, pp 403–406. <https://doi.org/10.2991/amsm-16.2016.90>
- Al-Humairi A, Zheng X, Ip RHL, El Masoud B (2016) Computed tomography image quality evaluation for pre-surgical dental implant site assessment using different exposure setting protocols: mandibular phantom study. In: 4th IIAE international conference on intelligent systems and image processing. The Institute of Industrial Applications Engineers, Japan, pp 300–305. <https://doi.org/10.12792/iciisip2016.053>
- Ali S, Ali A, Khan SA, Hussain S (2016) Sufficient sample size and power in multilevel ordinal logistic regression models. *Comput Math Methods Med* 2016:7329158. <https://doi.org/10.1155/2016/7329158>
- Almashraqi AA, Ahmed EA, Mohamed NS, Barnkgkei IH, Elsherbini NA, Halboub ES (2017) Evaluation of different low-dose multi-detector CT and cone beam CT protocols in maxillary sinus imaging: part I-an in vitro study. *Dentomaxillofac Radiol* 46:20160323. <https://doi.org/10.1259/dmfr.20160323>
- Anderson PJ, Yong R, Surman TL, Rajion ZA, Ranjitkar S (2014) Application of three-dimensional computed tomography in craniofacial clinical practice and research. *Aust Dent J* 59(Suppl 1):174–185. <https://doi.org/10.1111/adj.12154>
- Attard S, Castillo J, Zarb F (2018) Establishment of image quality for MRI of the knee joint using a list of anatomical criteria. *Radiography (lond)* 24:196–203. <https://doi.org/10.1016/j.radi.2018.01.008>
- Barrett HH, Myers KJ, Hoeschen C, Kupinski MA, Little MP (2015) Task-based measures of image quality and their relation to radiation dose and patient risk. *Phys Med Biol* 60:R1–R75. <https://doi.org/10.1088/0031-9155/60/2/r1>
- Båth M, Månsson LG (2007) Visual grading characteristics (VGC) analysis: a non-parametric rank-invariant statistical method for image quality evaluation. *Br J Radiol* 80:169–176. <https://doi.org/10.1259/bjr/35012658>
- Bauer DJ, Sterba SK (2011) Fitting multilevel models with ordinal outcomes: performance of alternative specifications and methods of estimation. *Psychol Methods* 16:373–390. <https://doi.org/10.1037/a0025813>
- Bidra AS (2017) Surgical safety checklist for dental implant and related surgeries. *J Prosthet Dent* 118:442–444. <https://doi.org/10.1016/j.prosdent.2017.02.019>
- Carter L, Farman AG, Geist J, Scarfe WC, Angelopoulos C, Nair MK, Hildebolt CF, Tyndall D, Shrout M (2008) American Academy of Oral and Maxillofacial Radiology executive opinion statement on performing and interpreting diagnostic cone beam computed tomography. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 106:561–562. <https://doi.org/10.1016/j.tripleo.2008.07.007>
- Christensen RHB. (2015) Analysis of ordinal data with cumulative link models—estimation with the R-package ordinal. [https://mran.microsoft.com/snapshot/2017-12-11/web/packages/ordinal/vignettes/clm\\_intro.pdf](https://mran.microsoft.com/snapshot/2017-12-11/web/packages/ordinal/vignettes/clm_intro.pdf)
- Christensen RHB (2019) A tutorial on fitting cumulative link mixed models with clmm2 from the ordinal package. [https://cran.r-project.org/web/packages/ordinal/vignettes/clmm2\\_tutorial.pdf](https://cran.r-project.org/web/packages/ordinal/vignettes/clmm2_tutorial.pdf)
- Dawood A, Brown J, Sauret-Jackson V, Purkayastha S (2012) Optimization of cone beam CT exposure for pre-surgical evaluation of the implant site. *Dentomaxillofac Radiol* 41:70–74. <https://doi.org/10.1259/dmfr/16421849>
- Dow JK, Endersby JW (2004) Multinomial probit and multinomial logit: a comparison of choice models for voting research. *Elect Stud* 23:107–122. [https://doi.org/10.1016/S0261-3794\(03\)00040-4](https://doi.org/10.1016/S0261-3794(03)00040-4)
- Drikvandi R, Noorian S (2019) Testing random effects in linear mixed-effects models with serially correlated errors. *Biom J* 61:802–812. <https://doi.org/10.1002/bimj.201700203>
- Favazza CP, Fetterly KA, Hangiandreu NJ, Leng S, Schueler BA (2015) Implementation of a channelized Hotelling observer model to assess image quality of x-ray angiography systems. *J Med Imaging (bellingham)* 2:015503. <https://doi.org/10.1117/1.jmi.2.1.015503>
- Froum S, Casanova L, Byrne S, Cho SC (2011) Risk assessment before extraction for immediate implant placement in the posterior mandible: a computerized tomographic scan study. *J Periodontol* 82:395–402. <https://doi.org/10.1902/jop.2010.100360>
- Goto TK, Nishida S, Nakamura Y, Tokumori K, Nakamura Y, Kobayashi K, Yoshida Y, Yoshiura K (2007) The accuracy of 3-dimensional magnetic resonance 3D vobe images of the mandible: an in vitro comparison of magnetic resonance imaging and computed tomography. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 103:550–559. <https://doi.org/10.1016/j.tripleo.2006.03.011>
- Harris D, Horner K, Gröndahl K, Jacobs R, Helmrot E, Benic GI, Bornstein MM, Dawood A, Quirynen M (2012) 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* 23:1243–1253. <https://doi.org/10.1111/j.1600-0501.2012.02441.x>
- Hedeker D, Gibbons RD (1994) A random-effects ordinal regression model for multilevel analysis. *Biometrics* 50:933–944. <https://doi.org/10.2307/2533433>
- Heetveld MJ, Raaymakers EL, van Walsum AD, Barei DP, Steller EP (2005) Observer assessment of femoral neck radiographs after reduction and dynamic hip screw fixation. *Arch Orthop Trauma Surg* 125:160–165. <https://doi.org/10.1007/s00402-004-0780-4>
- Hidalgo Rivas JA, Horner K, Thiruvencatachari B, Davies J, Theodorakou C (2015) Development of a low-dose protocol for cone beam CT examinations of the anterior maxilla in children. *Br J Radiol* 88:20150559. <https://doi.org/10.1259/bjr.20150559>
- Hidalgo-Rivas JA, Theodorakou C, Carmichael F, Murray B, Payne M, Horner K (2014) Use of cone beam CT in children and young people in three United Kingdom dental hospitals. *Int J Paediatr Dent* 24:336–348. <https://doi.org/10.1111/ipd.12076>
- Hofmann E, Schmid M, Lell M, Hirschfelder U (2014) Cone beam computed tomography and low-dose multislice computed tomography in orthodontics and dentistry. *J Orofac Orthop* 75:384–398. <https://doi.org/10.1007/s00056-014-0232-x>
- Jones A, Ansell C, Jerrom C, Honey ID (2015) Optimization of image quality and patient dose in radiographs of paediatric extremities using direct digital radiography. *Br J Radiol* 88:20140660. <https://doi.org/10.1259/bjr.20140660>
- Kadesjö N, Benchimol D, Falahat B, Näsström K, Shi X-Q (2015) Evaluation of the effective dose of cone beam CT and multislice CT for temporomandibular joint examinations at optimized exposure levels. *Dentomaxillofac Radiol* 44:20150041. <https://doi.org/10.1259/dmfr.20150041>
- Kim M-J, Lee S-S, Choi M, Yong HS, Lee C, Kim J-E, Heo M-S (2020) Developing evidence-based clinical imaging guidelines of justification for radiographic examination after dental implant installation. *BMC Med Imaging* 20:102. <https://doi.org/10.1186/s12880-020-00501-3>

- Kleinbaum DG, Kupper LL, Nizam A, Muller KE (2007) Applied regression analysis and other multivariable methods. Duxbury Press, Belmont
- Koizumi H, Sur J, Seki K, Nakajima K, Sano T, Okano T (2010) Effects of dose reduction on multi-detector computed tomographic images in evaluating the maxilla and mandible for pre-surgical implant planning: a cadaveric study. *Clin Oral Implants Res* 21:830–834. <https://doi.org/10.1111/j.1600-0501.2010.01925.x>
- Krantz J (2012) Experiencing sensation and perception. Pearson Education, Ltd., Upper Saddle River
- Kundel HL (2015) Visual search and lung nodule detection on CT scans. *Radiology* 274:14–16. <https://doi.org/10.1148/radiol.14142247>
- Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174. <https://doi.org/10.2307/2529310>
- Lee KC, Bamford A, Gardiner F, Agovino A, ter Horst B, Bishop J, Sitch A, Grover L, Logan A, Moiemens NS (2019) Investigating the intra- and inter-rater reliability of a panel of subjective and objective burn scar measurement tools. *Burns* 45:1311–1324. <https://doi.org/10.1016/j.burns.2019.02.002>
- Liang X, Jacobs R, Hassan B, Li L, Pauwels R, Corpas L, Couto Souza P, Martens W, Shahbazian M, Alonso A, Lambrichts I (2010) A comparative evaluation of cone beam computed tomography (CBCT) and multi-slice CT (MSCT) Part I. On subjective image quality. *Eur J Radiol* 75:265–269. <https://doi.org/10.1016/j.ejrad.2009.03.042>
- Likert R (1932) A technique for the measurement of attitudes. *Arch Psychol* 22:1–55
- Liu LC, Hedeker D (2006) A mixed-effects regression model for longitudinal multivariate ordinal data. *Biometrics* 62:261–268. <https://doi.org/10.1111/j.1541-0420.2005.00408.x>
- Lofthag-Hansen S, Thilander-Klang A, Gröndahl K (2011) Evaluation of subjective image quality in relation to diagnostic task for cone beam computed tomography with different fields of view. *Eur J Radiol* 80:483–488. <https://doi.org/10.1016/j.ejrad.2010.09.018>
- Månsson LG (2000) Methods for the evaluation of image quality: a review. *Radiat Prot Dosim* 90:89–99. <https://doi.org/10.1093/oxfordjournals.rpd.a033149>
- Metz CE (1989) Some practical issues of experimental design and data analysis in radiological ROC studies. *Invest Radiol* 24:234–245. <https://doi.org/10.1097/00004424-198903000-00012>
- Mraity H, England A, Hogg P (2014) Developing and validating a psychometric scale for image quality assessment. *Radiography* 20:306–311. <https://doi.org/10.1016/j.radi.2014.04.002>
- Nemtoi A, Czink C, Haba D, Gahleitner A (2013) Cone beam CT: a current overview of devices. *Dentomaxillofac Radiol* 42:20120443. <https://doi.org/10.1259/dmfr.20120443>
- Park H-N, Min C-K, Kim K-A, Koh K-J. (2019) Optimization of exposure parameters and relationship between subjective and technical image quality in cone-beam computed tomography. *Imaging Sci Dent* 49: 139–151. <https://doi.org/10.5624/isd.2019.49.2.139>
- Pittayapat P, Galiti D, Huang Y, Dreesen K, Schreurs M, Couto Souza P, Rubira-Bullen IRF, Westphalen FH, Pauwels R, Kalema G, Willems G, Jacobs R (2013) An in vitro comparison of subjective image quality of panoramic views acquired via 2D or 3D imaging. *Clin Oral Investig* 17:293–300. <https://doi.org/10.1007/s00784-012-0698-0>
- Powers D, Xie Y (2008) Statistical methods for categorical data analysis. Bingley, Emerald
- Prasad SR, Wittram C, Shepard JA, McLoud T, Rhea J (2002) Standard-dose and 50%-reduced-dose chest CT: comparing the effect on image quality. *AJR Am J Roentgenol* 179:461–465. <https://doi.org/10.2214/ajr.179.2.1790461>
- R Core Team (2020) R: a language and environment for statistical computing. [www.r-project.org](http://www.r-project.org).
- Rios HF, Borgnakke WS, Benavides E (2017) The use of cone-beam computed tomography in management of patients requiring dental implants: an American Academy of Periodontology best evidence review. *J Periodontol* 88:946–959. <https://doi.org/10.1902/jop.2017.160548>
- Rossmann K, Wiley BE (1970) The central problem in the study of radiographic image quality. *Radiology* 96:113–118. <https://doi.org/10.1148/96.1.113>
- Saffari SE, Löve Á, Fredrikson M, Smedby Ö (2015) Regression models for analyzing radiological visual grading studies – an empirical comparison. *BMC Med Imaging* 15:49. <https://doi.org/10.1186/s12880-015-0083-y>
- Sakata-Goto T, Takahashi K, Kiso H, Huang B, Tsukamoto H, Takemoto M, Hayashi T, Sugai M, Nakamura T, Yokota Y, Shimizu A, Slavkin H, Bessho K (2012) Id2 controls chondrogenesis acting downstream of BMP signaling during maxillary morphogenesis. *Bone* 50:69–78. <https://doi.org/10.1016/j.bone.2011.09.049>
- Schaefferkoetter JD, Yan J, Townsend DW, Conti M (2015) Initial assessment of image quality for low-dose PET: evaluation of lesion detectability. *Phys Med Biol* 60:5543–5556. <https://doi.org/10.1088/0031-9155/60/14/5543>
- SEDENTEXCT (2012) Radiation Protection No. 172—cone beam CT for dental and maxillofacial radiology (evidence-based guidelines). <https://ec.europa.eu/energy/sites/ener/files/documents/172.pdf>
- Shelley AM, Brunton P, Horner K (2011) Subjective image quality assessment of cross sectional imaging methods for the symphyseal region of the mandible prior to dental implant placement. *J Dent* 39:764–770. <https://doi.org/10.1016/j.jdent.2011.08.008>
- Smedby Ö, Fredrikson M (2010) Visual grading regression: analysing data from visual grading experiments with regression models. *Br J Radiol* 83:767–775. <https://doi.org/10.1259/bjr/35254923>
- Smedby Ö, Fredrikson M, De Geer J, Borgen L, Sandborg M (2013) Quantifying the potential for dose reduction with visual grading regression. *Br J Radiol* 86:31197714–31197714. <https://doi.org/10.1259/bjr/31197714>
- Spuur K (2019) A review of mammographic lesion localisation and work up imaging in Australia in the digital era. *Radiography* 25:385–391. <https://doi.org/10.1016/j.radi.2019.03.002>
- Sund P, Båth M, Kheddache S, Månsson LG (2004) Comparison of visual grading analysis and determination of detective quantum efficiency for evaluating system performance in digital chest radiography. *Eur Radiol* 14:48–58. <https://doi.org/10.1007/s00330-003-1971-z>
- Tanny L, Huang B, Naung NY, Currie G (2018) Non-orthodontic intervention and non-nutritive sucking behaviours: a literature review. *Kaohsiung J Med Sci* 34:215–222. <https://doi.org/10.1016/j.kjms.2018.01.006>
- Thornbury JR, Fryback DG, Patterson FE, Chiavarini RL (1978) Effect of screen/film combinations on diagnostic certainty: Hi-Plus/RPL versus Lanex/Ortho G in excretory urography. *AJR Am J Roentgenol* 130:83–87. <https://doi.org/10.2214/ajr.130.1.83>
- Vandenberghe B, Jacobs R, Yang J (2007) Diagnostic validity (or acuity) of 2D CCD versus 3D CBCT-images for assessing periodontal breakdown. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 104:395–401. <https://doi.org/10.1016/j.tripleo.2007.03.012>
- Zanca F, Van Ongeval C, Claus F, Jacobs J, Oyen R, Bosmans H (2012) Comparison of visual grading and free-response ROC analyses for assessment of image-processing algorithms in digital mammography. *Br J Radiol* 85:e1233–e1241. <https://doi.org/10.1259/bjr/22608279>
- Zarb F, McEntee MF, Rainford L (2015) Visual grading characteristics and ordinal regression analysis during optimisation of CT head examinations. *Insights Imaging* 6:393–401. <https://doi.org/10.1007/s13244-014-0374-9>

- Zheng X (2017) General equations for optimal selection of diagnostic image acquisition parameters in clinical X-ray imaging. *Radiol Phys Technol* 10:415–421. <https://doi.org/10.1007/s12194-017-0413-6>
- Zheng X, Kim M, Yang S (2016) Optimal kVp in chest computed radiography using visual grading scores: a comparison between visual grading characteristics and ordinal regression analysis. In: *Proceedings of the SPIE 9783, medical imaging 2016: physics of medical imaging*, 97836A. <https://doi.org/10.1117/12.2217414>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.