RESEARCH ARTICLE

Does anthropomorphic model design in *ex vivo* studies affect diagnostic accuracy for dental root fracture using CBCT?

^{1,2}Fedil Andraws Yalda, ³Rosalyn J Clarkson, ⁴Jonathan Davies, ⁵Peter G J Rout, ²Anita Sengupta and ²Keith Horner

¹College of Dentistry, Hawler Medical University, Erbil, Iraq; ²Division of Dentistry, School of Medical Science, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Oxford Road, Manchester M13 9PL, UK; ³Department of Radiology, Leeds Dental Institute, Clarendon Way, Leeds LS2 9LU, UK; ⁴Department of Dental Radiological Imaging, Guy's and St Thomas' NHS Foundation Trust, Floor 23, Tower Wing, Guy's Hospital, Great Maze Pond, London SE1 9RT, UK; ⁵Birmingham Dental Hospital, Birmingham Community Healthcare Foundation NHS Trust, 5 Mill Pool Way, Birmingham, B5 7EG, UK

Objectives: The evidence for diagnostic accuracy using cone beam computed tomography (CBCT) for dental applications depends heavily on *ex vivo* research, but there is little knowledge of whether the model used affects the diagnostic accuracy results. The objective of this study was to determine the impact of different designs of anthropomorphic models on diagnostic accuracy for the specific task of dental root fracture detection.

Methods: Horizontal or oblique root fracture was induced in 24 of 48 permanent maxillary incisors. The 48 teeth were scanned by CBCT using standard clinical exposure factors on five occasions, each with a different model design. Scans were viewed by five dental and maxillofacial radiologists, who each made a forced diagnosis of fracture or no fracture in each root and a judgment on root fracture using a five-point confidence scale. Sensitivity (Se), specificity (Sp) and areas under receiver operating characteristic (ROC) curve (Az) were calculated for each observer for each model.

Results: There were no significant differences between the diagnostic accuracy measurements recorded using different models. There were, however, numerous significant differences between observers using the same anthropomorphic model.

Conclusions: Despite the differences in X-ray attenuation between the five model designs, the results suggest that the anthropomorphic model does not affect the results of diagnostic accuracy studies on root fracture using this CBCT machine at standard clinical exposures. This provides some confidence in the previously published evidence. The interobserver diagnosis differences indicate that research using only two observers could provide misleading results.

Dentomaxillofacial Radiology (2020) **49**, 20200093. doi: [10.1259/dmfr.20200093](https://doi.org/10.1259/dmfr.20200093)

Cite this article as: Andraws Yalda F, Clarkson RJ, Davies J, Rout PGJ, Sengupta A, Horner K. Does anthropomorphic model design in *ex vivo* studies affect diagnostic accuracy for dental root fracture using CBCT?. *Dentomaxillofac Radiol* 2020; **49**: 20200093.

Keywords: Cone-Beam Computed Tomography; Diagnosis; Tooth Fractures; Research Design; Phantoms; Imaging

Introduction

Trauma to maxillary anterior teeth is common and may result in root fracture. As these are hidden from direct visual inspection, imaging using X-rays is always

required when fracture is suspected.¹ In the last decade, conventional radiographical examinations have been supplemented by cone beam computed tomography (CBCT). The systematic review by Salinero et al. suggested that diagnostic accuracies using CBCT can be very high for fracture detection in teeth without

Correspondence to: Dr Fedil Andraws Yalda, E-mail: dr-fedil@hotmail.com

Received 10 March 2020; revised 03 May 2020; accepted 21 May 2020

root canal restorations, reporting a pooled sensitivity of 0.82, a pooled specificity of 0.88, and a diagnostic odds ratio of 39.05 for teeth without metal artifact.² Their review found that the primary research evidence was overwhelmingly based on *ex vivo* studies evaluating artificial fractures in extracted teeth. A systematic review of *in vivo* studies was conducted by Long *et al*³ but, however, the studies in their review included endodontically treated teeth and predominantly vertical fractures, so the findings do not relate well to the typical situation of tooth trauma in children and young people. Furthermore, *in vivo* studies are few in number and very challenging because a reference standard diagnosis of root fracture requires either extensive surgical exposure of the root or extraction of the tooth, both of which are unacceptable ethically. Despite the paucity of evidence from clinical studies, various guideline documents, recently collated as part of a systematic review, have recommended that localized CBCT may be used in selected cases for detection of root fracture because of better diagnostic accuracy.⁴

Because of the reliance on *ex vivo* studies as the evidence for diagnostic accuracy using CBCT for root fractures, and also for numerous other dental diagnostic tasks, it is important to be assured that the research models used adequately reproduce the clinical situation, in other words that they are anthropomorphic, as recently reviewed fully by the main author.⁵ Citing only a few examples here, some studies such as those by Mohammadpour *et al*⁶ and Moudi *et al*⁷ have simply used teeth set in acrylic blocks, while others have used limited amounts of dried bone material such as isolated mandibles or hemi-mandibles.^{8,9} Other studies have used non-human or artificial bone.^{10–12} Some researchers have included no soft tissue substitute in their models,^{8,9} whereas many have added insufficient thicknesses of soft tissue equivalent materials, such as a few layers of dental wax.^{13–15} Jakobson *et al.* employed only 1.5 mm bovine muscle as the soft tissue substitute.¹⁶ In other studies, a hollow acrylic cylinder was used to mimic soft tissues of the face and neck, although the reason for such a design is unexplained.^{17,18} Only a few studies seem to have paid attention to reproducing the attenuation characteristics of the human head, employing complete jaws, and sometimes cervical vertebrae, all immersed in a water bath for soft tissue simulation.^{18–20}

Why should the research model design matter? X-ray attenuation is dependent on the quantity and types of material through which the X-ray beam passes. *In vivo*, the bone and soft tissues of the face and neck absorb and scatter radiation. In *ex vivo* studies, upon which much of the evidence is based, the model design could significantly change the quantity of radiation reaching the image receptor and the amount of scatter. Thus, image quality would be affected and this, in turn, might significantly influence diagnostic accuracies achieved by clinicians using CBCT. This assumption is based on a fundamental understanding of X-ray physics, but the

impact on diagnostic accuracy lacks empirical evidence. Without such information, some doubt must exist about the data on diagnostic accuracy and the strength of the recommendations made in current clinical guidelines about the use of CBCT in root fracture diagnosis. Furthermore, previous systematic reviews related to diagnostic accuracy for root fractures using CBCT have given little or no attention to this potential source of performance bias. While the context considered here has been root fracture diagnosis, the same issue is of relevance to other *ex vivo* research on the diagnostic accuracy using CBCT, such as for detection of caries, resorptions, and periapical bone defects.

The aim of this study was, therefore, to determine the impact of different designs of anthropomorphic models on diagnostic accuracy of observers using CBCT for dental root fracture detection.

Methods and materials

Ethical aspects

As the teeth were sourced from outside the UK, NHS ethics approval was not applicable. Teeth were stored before, during the study when not in use, and after completion of the research following relevant national legal requirements. The study protocol was reviewed and approved by The University of Manchester Research Ethics Committee (Ref: 2018-4119-5965). Informed consent was obtained from the five observers for their involvement.

Teeth

The teeth used in the study were sourced from patients attending for dental extractions as part of normal clinical care at the College of Dentistry, Hawler Medical University, Erbil, Iraq. Collection was made with patient consent and followed the local and national ethical requirements. Teeth were stored in individual laboratory glass containers containing 10 ml of neutral buffered 10% formalin solution (Sigma-Aldrich Company Ltd., Dorset, UK).

To determine sample size of teeth for the study, diagnostic accuracy was defined as the area under a receiver operating characteristic (ROC) curve. The required sample size for the comparison of the areas under two ROC curves (derived from the same cases) was carried out using MedCalc Statistical Software v.18.10.2 (MedCalc Software bvba, Ostend, Belgium; <https://www.medcalc.org>; 2018). To perform the calculation, the selected α -level (two-sided) was 0.05 and the β -level was 0.20 (power of 80%). The hypothesized areas of the two ROC curves were set at 0.93 and 0.83. The area under ROC curve one was obtained from the mean value of two previous studies that used the same CBCT scanner as the one planned to be used in the current study.^{18,21} A difference of 0.10 in area was assumed for the second ROC curve. The correlation in positive and negative

groups was derived from the mean κ values of interobserver repeatability obtained in previous studies.^{18,21–23} The ratio of sample sizes in negative/positive groups was chosen to be 1:1. The total sample size calculated using these assumptions was 48 (two groups of 24 teeth; one group with and the other without root fracture).

55 maxillary permanent incisor teeth were initially collected (28 central incisors and 27 lateral incisors), from which 48 were randomly selected for the study (<https://www.random.org/integers/>). These were, in turn, randomly divided into the required two groups of 24 teeth (root fracture and no root fracture). The spare teeth were retained as possible reserves, if required.

The possibility of pre-existing root fractures within the sample was assessed by performing high-resolution micro-CT scans using the SkyScan 1072 (SkyScan, Kontich, Belgium). Each tooth was positioned on the specimen stage and scanned using standard settings at a resolution of 11.24 μm , tube voltage of 100 kV, tube current of 98 mA and with an exposure time of 1900 ms. Datasets were processed using NRecon software (V.1.6.4.7, SkyScan). Once processed, one person (FY) viewed microtomographic cross-sections generated using DataViewer v.1.4.4.0, SkyScan software. This information also provided the reference standard for the 24 teeth not selected for fracture.

Tooth preparation and fracture induction

To induce root fracture in the horizontal or oblique planes, each tooth was individually mounted in 15 mm high acrylic resin blocks into which a cylindrical hole had been drilled. Each tooth was retained in the hole using a layer of dental sticky wax (DeguDent GmbH, Hanau-Wolfgang, Germany). The acrylic resin block was fixed in an impact machine (Monsanto tensometer plastic impact machine, Shakopee, USA). In use, the pendulum was raised at an angle of 145° to the horizontal. Weights at the free end of the pendulum were initially added or removed by a process of trial and error, finally settling on weights of 114 g and 292 g for fracturing lateral and central incisors, respectively. The fragments of fractured roots were reassembled and fixed using cyanoacrylate glue (Loctite® super glue, Henkel Corp., USA), following the methods of Wenzel *et al* and Jones *et al*.^{18,24}

After fracture and reassembling of fragments, high-resolution micro-CT scans were repeated using the same methodology as described above. This was done to ensure that no fractures were present in the teeth other than that induced for the study. This information provided the reference standard for the 24 fractured teeth.

Anthropomorphic model design

Five anthropomorphic model designs into which teeth could be positioned were used in the study:

- Model 1: Curved acrylic resin block.

- Model 2: Dried bone, consisting of all parts of the facial skeleton but without the superior half of the cranial vault (positioning teeth in anterior maxillary tooth sockets).
- Model 3: As Model 2, but with limited soft tissue equivalent material (5 mm wax) overlying the anterior maxillary alveolus.
- Model 4: As Model 2, but after pre-soaking the bony material in water and with immersion in a head-shaped water bath to give full soft tissue equivalent.
- Model 5: As Model 4, but with inclusion of cervical vertebrae.

These are illustrated in [Figure 1](#).

The acrylic block used as Model 1 was manufactured using conventional methods for constructing a polymethylmethacrylate (acrylic) denture. A preliminary model was manufactured in red wax (Metrodent Ltd, Paddock, Huddersfield, UK), corresponding to the shape of the anterior part of the maxillary alveolus using the author's judgment. This was then used to manufacture a final model in heat-cured acrylic (Metrodent Ltd, Paddock, Huddersfield, UK). The dimensions of the block were 25×25×110 mm curved length.

For Models 2–5, inclusive, a human skull from an historical collection was required. A small adult skull, most likely to be female, was selected because the six maxillary anterior teeth had been lost, leaving empty sockets, permitting alternative teeth to be inserted and removed for the study. The skull was complete apart from the upper part of the cranial vault. For Model 5, human cervical spine vertebrae (C1 to C5 inclusive) were added.

For Model 3, a 5 mm thickness of utility wax (Metrodent Ltd, Paddock, Huddersfield, UK) was applied to the labial surface of the anterior maxilla to cover the six anterior tooth sockets and extended to the first premolar regions bilaterally and, in the vertical direction, from the crestal alveolar bone to the anterior nasal spine. This thickness of wax was chosen as it reflected the models used in some previous diagnostic accuracy studies.^{14,15}

For Models 4 and 5, prior to radiographical imaging the skeletal material was soaked for 90 min in warm water into which hand dishwashing liquid had been added to displace air that would otherwise alter the X-ray attenuation properties. For Models 4 and 5, the skull was imaged while positioned in a water bath, using the container designed and described by Shelley *et al*, after minor modification to extend its height to allow the maxilla to be completely submerged in water.²⁵

CBCT scanning

A 3D Accuitomo 170 scanner (J. Morita, Kyoto, Japan) was used, operated at 90 kV, 5.0 mA and a full 360° rotation for 17.5 s, with a field of view (FOV) of 6×6 cm, using voxel size of 0.125 mm. Groups of six teeth were scanned at once, in keeping with the natural situation of six human anterior teeth in the dentate maxilla. Thus, for each phantom design, eight combinations of

Figure 1 The five models used in the study. (a) Model 1, showing teeth mounted in the acrylic resin block. (b) Model 1, showing the acrylic block positioned in the CBCT machine. (c) Model 2, mounted on a tripod and set up in the CBCT machine. (d) Model 3. (e) Water bath containing skeletal material, mounted on a tripod, used for Models 4 and 5; the latter included cervical vertebrae.

teeth were required to image all 48 teeth, resulting in 40 CBCT datasets. Fractured and non-fractured teeth were randomly allocated to these eight combinations. Eight CBCT datasets (20%) were randomly selected from the 40 and duplicated to permit an assessment of intraobserver repeatability. The order of the resulting 48 CBCT datasets was randomized and each was allocated a study number.

The CBCT image datasets were exported from the acquisition computer with the native viewing software (i-Dixel, One Volume Viewer, Morita, Japan) to a Dell laptop (Inspiron 17 7000 Series 2-in-1-7773, Dell Inc., TX, USA) 17.3-inch FHD, IPS Truelife LED-Backlit Touch Display with a screen resolution of 1920×1080 pixels, luminance of 330cd/m², and luminance contrast ratio of 1538:1. These parameters fulfilled

the requirements for monitors used for clinical CBCT viewing, as previously defined.²⁶

CBCT assessments

Five observers, all registered specialists in Dental and Maxillofacial Radiology (DMFR), assessed the 48 CBCT datasets independently and were blinded to the reference standard diagnosis. Each was allocated a study number (observer 1, observer 2, etc.). In terms of experience in DMFR, this ranged from 7 to over 30 years with experience of using CBCT between 6 and 12 years. All had previous experience with the i-Dixel viewing software. All assessments were made on the same laptop PC described above. Observers had complete freedom to explore the image volume, perform reconstructions and to adjust

image contrast and brightness as they wished when performing the assessments. Printed instructions were provided to the observers to explain the assessment tasks. This included a brief scenario to provide a clinical context:

“Each scan is of a patient who has recently undergone trauma to the anterior maxillary teeth. You have examined the patient and the teeth are loosened. You have taken conventional radiographs, but you are still concerned that there may be undetected root fractures and feel that a localized CBCT is justified”.

Observers were asked to evaluate each tooth root on each of the CBCT scans in two ways. First, they were asked to make a clear decision on either the presence or absence of a root fracture. Second, the observers recorded their judgment on the presence of root fracture using a five-point confidence scale: 1: definitely absent; 2: probably absent; 3: not sure; 4: probably present; and 5: definitely present. All data were pseudonymized to provide as much anonymity as possible to the observers.

Statistical analysis

The null hypothesis tested was that there were no significant differences in diagnostic accuracy of observers using CBCT for the identification of root fractures in maxillary anterior teeth according to anthropomorphic model design. The diagnostic accuracy of each observer, for each of the five models, was measured by comparison with the true diagnosis of root fracture presence or absence (reference standard). Sensitivity (Se), specificity (Sp), accuracy (Ac), positive predictive value (PPV), and negative predictive value (NPV) were calculated for each observer and each model, using their yes/no answers for the presence of root fracture. ROC curve areas were calculated based on the five-point confidence scale for presence of root fracture. Intraobserver and interobserver reproducibility (agreement) for diagnosis of root fracture was measured using κ for the yes/no answers and weighted κ (κ_w) for 5-point scale answers) statistics with the scale for strength of agreement described by Altman: ≥ 0.20 poor, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good and 0.81–1.0 very good.²⁷

Statistical comparison of diagnostic accuracies was made by comparing the areas under different ROC curves (Az) using the method of DeLong *et al*, which allows for multiple comparisons between subjects when testing for significance.²⁸ To test the null hypothesis, comparisons of Az values achieved by each observer for the five models were performed using the method of Hanley and McNeil.²⁹ To perform the statistical calculations, MedCalc Statistical Software v.18.10.2 (MedCalc Software bvba, Ostend, Belgium; <https://www.medcalc.org>; 2018) was employed. The significance threshold was set at $p = 0.05$.

Results

For diagnosis of root fracture using yes/no answers, there was very good intraobserver agreement for observers 1 ($\kappa = 0.96$), 3 ($\kappa = 0.83$), 4 ($\kappa = 0.83$) and 5 ($\kappa = 0.83$) while good agreement was achieved by observer 2 ($\kappa = 0.79$). Intraobserver agreement for the five-point confidence scale answers was very good for observer 1 ($\kappa_w = 0.84$) while the other four showed good intraobserver agreement (observer 2, $\kappa_w = 0.62$; observer 3, $\kappa_w = 0.72$; observer 4, $\kappa_w = 0.68$; observer 5, $\kappa_w = 0.74$). Interobserver agreement for the diagnosis of fracture (yes/no answers) varied from moderate-to-very good (median $\kappa = 0.75$; minimum $\kappa = 0.53$ to a maximum $\kappa = 1.00$). Interobserver agreement for the five-point confidence scale ranged from moderate-to-good (median $\kappa_w = 0.59$; minimum $\kappa_w = 0.40$ to a maximum $\kappa_w = 0.76$). There was no discernible difference in agreement levels between the models.

The diagnostic accuracy results for each observer, for each model, are presented in Table 1. Sensitivity and specificity values were derived from the observers' yes/no answers for the presence of root fracture. ROC curve areas (Az) were derived from the five-point scale answers. For each model, variation between observers in performance was evident, although sensitivities were always lower than specificities, regardless of the observer or the model. The mean Az $\pm$ s.d. of all five observers for each model (Model 1–Model 5, inclusive) were 83.0% $\pm$ 0.07, 86.7% $\pm$ 0.06, 84.6% $\pm$ 0.06, 83.7% $\pm$ 0.06, and 83.6% $\pm$ 0.06, respectively.

The Az for all observers, for all five models, were significantly different to 0.5 ($p < 0.001$). Comparisons of the Az values of observers for each model are shown in Table 2. Statistically significant differences were seen in 16 out of 50 potential interobserver, intramodel, comparisons, but observers 3 and 4 were each involved in four of these and observer 5 was involved in eight significant differences. Looking in detail at Tables 1 and 2 in parallel, observers 3 and 5 frequently had significantly lower diagnostic accuracies than the other three observers. While there were these statistically significant differences in diagnostic accuracies *between* observers for each model, when the five models were compared for each observer (*e.g.*, Model 1/Observer 1 versus Model 2/Observer 1, or Model 2/Observer 4 versus Model 3/Observer 4), no statistically significant differences were seen. Thus, the model design had no effect on the diagnostic accuracy of any observer and the null hypothesis was accepted.

Discussion

Root fracture was selected for the investigation because it has been the subject of numerous previous diagnostic accuracy studies with a wide range of anthropomorphic model designs. Root fractures can be induced for research purposes in a manner which reproduces

Table 1 The diagnostic accuracy results for the five observers for the five models. In each case, values are given as a percentage, to one decimal place, with 95% CIs in parentheses

		Observer				
		1	2	3	4	5
Model 1	Se	75.0 (53.3–90.2)	75.0 (53.3–90.2)	66.7 (44.7–84.4)	83.3 (62.6–95.3)	54.2 (32.8–74.5)
	Sp	100 (85.8–100)	100 (85.8–100)	79.2 (57.9–92.9)	91.7 (73.0–98.9)	100 (85.8–100)
	Az	85.4 (72.2–93.9)	87.3 (72.2–93.9)	72.7 (57.9–84.5)	89.2 (76.9–96.3)	80.5 (65.5–90.5)
Model 2	Se	75.0 (53.3–90.2)	75.0 (53.3–90.2)	70.8 (48.9–87.4)	87.5 (67.6–97.3)	58.3 (36.6–77.9)
	Sp	100 (85.8–100)	100 (85.8–100)	75.0 (53.3–90.2)	95.8 (78.9–99.9)	100 (85.8–100)
	Az	90.6 (78.7–97.1)	86.5 (73.5–94.6)	79.8 (65.7–90.0)	94.4 (83.6–99.0)	80.5 (65.5–90.5)
Model 3	Se	70.8 (48.9–87.4)	66.7 (44.7–84.4)	70.8 (48.9–87.4)	83.3 (62.6–95.3)	54.2 (32.8–74.5)
	Sp	95.8 (78.9–99.9)	95.8 (78.9–99.9)	91.7 (73.0–99.0)	95.8 (78.9–99.9)	100 (85.8–100)
	Az	81.5 (66.7–91.2)	88.5 (76.0–95.9)	83.3 (69.8–92.5)	93.1 (81.9–98.4)	76.6 (62.1–87.6)
Model 4	Se	70.8 (48.9–87.4)	66.7 (44.7–84.4)	66.7 (44.7–84.4)	83.4 (62.6–95.3)	50.0 (29.1–70.9)
	Sp	100 (85.8–100)	95.8 (78.9–99.9)	95.8 (78.9–99.9)	100 (85.8–100)	100 (85.8–100)
	Az	89.9 (77.8–96.7)	78.1 (63.8–88.8)	81.9 (68.2–91.5)	89.4 (77.1–96.4)	79.2 (65.0–89.5)
Model 5	Se	75.0 (53.3–90.2)	75.0 (53.3–90.2)	66.7 (44.7–84.4)	79.2 (57.9–92.9)	50.0 (29.1–70.9)
	Sp	100 (85.8–100)	95.8 (78.9–99.9)	95.8 (78.9–99.9)	95.8 (78.9–99.9)	95.8 (78.9–99.9)
	Az	92.2 (80.7–98.0)	85.2 (72.0–93.8)	77.8 (63.5–88.5)	90.2 (78.1–96.9)	72.6 (57.8–84.4)

Az, area under the ROC curve; Se, sensitivity; Sp, specificity.

reasonably well the appearance of “real,” minimally displaced, fractures before any resorption of margins has occurred. This is in contrast to artificial resorption lesions or periapical bone cavities, which are not easy to reproduce artificially in a research model.

The hypothesis underlying the current study was that the model used for *ex vivo* radiological studies using CBCT should be truly anthropomorphic, mimicking the structure and content of the human as genuinely as possible, including the X-ray attenuation properties. The clear result was that the model design had no significant impact on diagnostic accuracy for the five observers, even when comparing an acrylic block to support the teeth (Model 1) with a complete jaws and cervical spine in a water bath (Model 5). This result was unexpected bearing in mind the impact of the attenuation effects of hard and soft tissues. There was not even a suggestion of a trend for different diagnostic accuracy according to model design. Why might this have happened, when the impact of less or inappropriate amounts of hard and soft tissues on X-ray attenuation cannot be doubted? Logically, the highest level of contrast and the best contrast-to-noise ratio would be expected with Model 1, with a gradual reduction in these as more hard and soft tissue equivalent material is added. Thus, Model 5, the model most like a real patient, would be expected to have the highest attenuation of radiation and a reduction in image quality. The first possible reason why this was not observed in the study is that such effects, while present, are simply insufficient to alter diagnostic performance. Second, the freedom for the observers to

adjust image contrast and brightness might have been enough to overcome any effect on image quality of the phantom design. A third option might be related to the level of X-ray exposure. The settings used in the study were consistent with the manufacturer’s recommended exposure setting of 90 kV, full rotation and 4 to 6 mA. Image quality and X-ray exposure are normally inversely related. It is tempting to hypothesize that the manufacturer’s recommended level of exposure is higher than needed and that, regardless of the amount of hard or soft tissue equivalent material present, there would be adequate image quality. Therefore, if this were true, the impact of phantom design in *ex vivo* studies would only be apparent at lower exposure settings. Regardless of this, the study clearly showed that with this CBCT equipment the anthropomorphic model design did not significantly alter diagnostic performance. This provides some reassurance about the results of previous research in the literature, although this should not lead to complacency when the same may not be the case for other CBCT scanners with different detector systems and image quality capabilities.

While model design had no significant impact on diagnostic accuracy, there were significant differences between observers. As noted above, observers 3 and 5 recorded significantly lower diagnostic accuracies in multiple instances, with the former tending toward a lower specificity, while the latter had a tendency to lower sensitivity. Such variation in diagnostic performance is widely recognized, but in the current study it was not obviously related to experience as a radiologist, although

Table 2 Statistical comparison of areas under the ROC curves (Az) of the five observers for the five models, using the test of DeLong *et al.* (1988), showing *p*-values in the Table cells. Statistically significant differences (*p* < 0.05) are highlighted in bold text

Model 1				
Observer	2	3	4	5
1	0.56	<0.05	0.29	0.41
2		<0.05	0.71	0.23
3			<0.01	0.18
4				0.13
Model 2				
Observer	2	3	4	5
1	0.43	<0.05	0.40	<0.05
2		0.25	0.13	0.19
3			<0.05	0.89
4				0.01
Model 3				
Observer	2	3	4	5
1	0.28	0.73	0.09	0.40
2		0.33	0.34	<0.05
3			<0.05	0.11
4				<0.01
Model 4				
Observer	2	3	4	5
1	0.05	0.11	0.91	<0.05
2		0.56	0.12	0.88
3			0.14	0.55
4				0.05
Model 5				
Observer	2	3	4	5
1	0.10	0.01	0.56	<0.01
2		0.19	0.30	<0.05
3			0.01	0.31
4				<0.01

to present more information on this here would risk loss of anonymity of the observers. The influence of experience on diagnostic performance was not an objective of this study, but further research in this area would be interesting. Lee *et al.* (2013) and Waite *et al.* (2017) found that fatigue, workload, interpretative speed, and distractions are among the confounders for interpretative errors while reporting on a radiograph.^{30,31} These confounders were among many other factors, so-called cognitive biases, which have been emphasized by most studies about diagnostic errors in radiology. Satisfaction of search bias is the tendency to conclude the observation once an abnormality has been found.³² Hindsight bias or recall is a tendency to overestimate the probability of a particular diagnosis when the correct diagnosis is already known and this happens mostly with senior radiologists than with trainees.³² Furthermore, differences in decision-making may have an effect on proper diagnosis such as past experience (making errors in the past), the mood on the day, fatigue, and the attitude to risk of making errors.

There is no consensus on the appropriate number of observers required in research studies, which perhaps depends on the task studied and the aims of the research.³³ Overall, the results of the current study provide evidence that studies with small numbers of observers may provide data that could be misleading. For example, if only two observers had been used in the current study, the results would have been very different if the pairing had been observers 3 and 5 or observers 1 and 4. It is worth noting, therefore, that a recent review as part of a doctoral thesis⁵ identified 59 diagnostic accuracy studies on root fracture using CBCT of which 13 included only two observers, some of which have been cited here, and one study that reported findings from a single observer.^{8,9,12,13,15,16,34}

While the purpose of the current study was focused on anthropomorphic model design, it is worth comparing our findings for diagnostic accuracy with previous research. Four previous studies have used a Morita CBCT scanner in the context of horizontal root fracture diagnosis,^{18,21–23} with two of these using the Accuitomo 170 model used in the current study.^{21,23} All of these reported higher sensitivities (all >91%) than we have reported here. Apart from Kamburoglu *et al.*,²¹ it is obvious that the FOV and exposure parameters/technical specifications used in these studies were different than those used in the current study, despite the use of the same scanner. The slightly larger FOV used in our study changes the default voxel dimensions on the 3D Accuitomo 170 to 0.125 mm, compared with the 0.08 mm that is used for a 4×4 cm FOV. This should have some impact on image quality. The 6×6 cm FOV used in this study was chosen for two reasons. First, it allowed imaging of the entire anterior maxillary segment, such as might be appropriate in cases of dental trauma to several anterior teeth. In addition, the voxel size of 0.125 mm is still within the definition of “high resolution” used in the European guidelines³⁵ and also more typical of the voxel sizes of CBCT equipment available on the market.³⁶ X-ray exposure settings for the current study were based on the manufacturer’s recommended levels, but it seems logical to expect that the setting chosen would have an impact on image quality that could change achievable diagnostic accuracy. Kamburoglu *et al.* used exactly the same exposure factors as this study, including FOV and voxel size, but used an anthropomorphic model which corresponds to our Model 3, applying a layer of wax to the labial surface of the anterior maxilla to reproduce soft tissue.²¹

No *ex vivo* research model can completely reproduce the clinical situation, which suffers from the unpredictable impact of patient movement and artefact-producing high attenuation material, usually dental restorations. Both of these are likely to lead to a reduction in image quality that might translate into lower diagnostic accuracy. As far as we have found, the effect of root filling materials and posts on diagnostic accuracy using CBCT have only been studied in the context of vertical root

fracture. In that situation diagnostic accuracy falls.^{2,3,37} It would be interesting to see whether this is also the case for the horizontal and oblique root fractures usually seen in acute trauma. Apart from the design of the model used in *ex vivo* research, there are numerous other factors which might affect diagnostic accuracies including the CBCT equipment model, the exposure settings (kV and mAs), exactly what is presented to observers (e.g., complete dataset or selected images) and the viewing conditions. In the specific context of root fractures, separation of fragments will have an impact on diagnostic accuracy. For the purpose of standardizing *ex vivo* diagnostic accuracy research, it would be beneficial to specify standard methods to improve the chance of being able to compare different studies in a meaningful way.

The best model is a real patient, but that poses challenges in obtaining a valid reference standard. Using cadavers for research can remove the concern about manufacturing an anthropomorphic model and allow a reference standard to be obtained. There are, however, challenging ethical aspects and restrictions on use of tissues from the deceased for research in some countries. Thus, it seems likely that we will continue to see anthropomorphic models used in dental radiological research in the future.

REFERENCES

1. Diangelis AJ, Andreasen JO, Ebeleseder KA, Kenny DJ, Trope M, Sigurdsson A, et al. Guidelines for the management of traumatic dental injuries: 1. fractures and Luxations of permanent teeth. *Pediatr Dent* 2017; **39**: 401–11.
2. Salineiro FCS, Kobayashi-Velasco S, Braga MM, Cavalcanti MGP. Radiographic diagnosis of root fractures: a systematic review, meta-analyses and sources of heterogeneity. *Dentomaxillofac Radiol* 2017; **46**: 20170400. doi: <https://doi.org/10.1259/dmfr.20170400>
3. Long H, Zhou Y, Ye N, Liao L, Jian F, Wang Y, et al. Diagnostic accuracy of CBCT for tooth fractures: a meta-analysis. *J Dent* 2014; **42**: 240–8. doi: <https://doi.org/10.1016/j.jdent.2013.11.024>
4. Horner K, Barry S, Dave M, Dixon C, Littlewood A, Pang CL, et al. Diagnostic efficacy of cone beam computed tomography in paediatric dentistry: a systematic review. *Eur Arch Paediatr Dent* Epub ahead of print. 2019 Dec 19 Erratum in: *Eur Arch Paediatr Dent*. 2020 Feb 3..
5. Andraws Yalda F. Aspects of dental and maxillofacial cone beam computed tomography (CBCT). *PhD thesis, University of Manchester* 2019;.
6. Mohammadpour M, Bakhshalian N, Shahab S, Sadeghi S, Ataee M, Sarikhani S. Effect of titanium and stainless steel posts in detection of vertical root fractures using NewTom v9 cone beam computed tomography system. *Imaging Sci Dent* 2014; **44**: 89–94. doi: <https://doi.org/10.5624/isd.2014.44.2.89>
7. Moudi E, Haghaniifar S, Madani Z, Bijani A, Nabavi ZS. The effect of metal artifacts on the identification of vertical root fractures using different fields of view in cone-beam computed tomography. *Imaging Sci Dent* 2015; **45**: 147–51. doi: <https://doi.org/10.5624/isd.2015.45.3.147>
8. Abdinian M, Razavian H, Jenabi N. In vitro comparison of cone beam computed tomography with digital periapical radiography for detection of vertical root fracture in posterior teeth. *J Dent* 2016; **17**: 84–90.
9. Elsaltani MH, Farid MM, Eldin Ashmawy MS. Detection of simulated vertical root fractures: which cone-beam computed tomographic system is the most accurate? *J Endod* 2016; **42**: 972–7. doi: <https://doi.org/10.1016/j.joen.2016.03.013>
10. Bechara B, Alex McMahan C, Moore WS, Noujeim M, Teixeira FB, Geha H. Cone beam CT scans with and without artefact reduction in root fracture detection of endodontically treated teeth. *Dentomaxillofac Radiol* 2013; **42**: 20120245. doi: <https://doi.org/10.1259/dmfr.20120245>
11. Parrone MT, Bechara B. Deahl ST 2nd, Ruparel Nb, Katkar R, Noujeim M. Cone beam computed tomography image optimization to detect root fractures in endodontically treated teeth: an in vitro (phantom) study. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2017; **123**: 613–20.
12. Taramsari M, Kajan ZD, Bashirzadeh P, Salamat F. Comparison of high-resolution and standard zoom imaging modes in cone beam computed tomography for detection of longitudinal root fracture: an in vitro study. *Imaging Sci Dent* 2013; **43**: 171–7. doi: <https://doi.org/10.5624/isd.2013.43.3.171>
13. Hassan B, Metska ME, Ozok AR, van der Stelt P, Wesselink PR. Comparison of five cone beam computed tomography systems for the detection of vertical root fractures. *J Endod* 2010; **36**: 126–9. doi: <https://doi.org/10.1016/j.joen.2009.09.013>
14. de Rezende Barbosa GL, Sousa Melo SL, Alencar PN, Nascimento MC, Almeida SM. Performance of an artefact reduction algorithm in the diagnosis of in vitro vertical root fracture in four different root filling conditions on CBCT images. *Int Endod J* 2016; **49**: 500–8.
15. Pinto MGO, Rabelo KA, Sousa Melo SL, Campos PSF, Oliveira LSAF, Bento PM, et al. Influence of exposure parameters on the detection of simulated root fractures in the presence of various intracanal materials. *Int Endod J* 2017; **50**: 586–94. doi: <https://doi.org/10.1111/iej.12655>

Conclusion

Anthropomorphic model design had no impact on the performance of observers using CBCT for root fracture diagnosis under the circumstances of this study. This suggests that concern about the validity of results from previous studies using models which were not anthropomorphic may be unfounded, at least for the diagnostic task considered here and the CBCT equipment used. Nonetheless, it seems appropriate to aspire to making *ex vivo* models in diagnostic accuracy studies as anthropomorphic as practicable; this might be of greater significance at lower X-ray exposures such as might be used in research on optimisation of exposures. Finally, there were significant differences between observers in diagnostic accuracies, highlighting that studies with only two raters might give misleading data.

Funding

Funding for this research was provided by the first author through their PhD fees. The authors declare no financial interests of relevance.

16. Jakobson SJM, Westphalen VPD, Silva Neto UX, Fariniuk LF, Schroeder AGD, Carneiro E. The influence of metallic posts in the detection of vertical root fractures using different imaging examinations. *Dentomaxillofac Radiol* 2014; **43**: 20130287. doi: <https://doi.org/10.1259/dmfr.20130287>
17. Patel S, Brady E, Wilson R, Brown J, Mannocci F. The detection of vertical root fractures in root filled teeth with periapical radiographs and CBCT scans. *Int Endod J* 2013; **46**: 1140–52. doi: <https://doi.org/10.1111/iej.12109>
18. Jones D, Mannocci F, Andiappan M, Brown J, Patel S. The Effect of Alteration of the Exposure Parameters of a Cone-beam Computed Tomographic Scan on the Diagnosis of Simulated Horizontal Root Fractures. *J Endod* 2015; **41**: 520–5. doi: <https://doi.org/10.1016/j.joen.2014.11.022>
19. Tangari-Meira R, Vancetto JR, Dovigo LN, Tosoni GM. Influence of Tube Current Settings on Diagnostic Detection of Root Fractures Using Cone-beam Computed Tomography: An In Vitro Study. *J Endod* 2017; **43**: 1701–5. doi: <https://doi.org/10.1016/j.joen.2017.05.008>
20. Menezes RFde, Araújo NCde, Santa Rosa JMC, Carneiro VSM, Santos Neto APdos, Costa V, et al. Detection of vertical root fractures in endodontically treated teeth in the absence and in the presence of metal post by cone-beam computed tomography. *BMC Oral Health* 2016; **16**: 48. doi: <https://doi.org/10.1186/s12903-016-0207-y>
21. Kamburoğlu K, Önder B, Murat S, Avsever H, Yüksel S, Paksoy CS. Radiographic detection of artificially created horizontal root fracture using different cone beam CT units with small fields of view. *Dentomaxillofac Radiol* 2013; **42**: 20120261. doi: <https://doi.org/10.1259/dmfr.20120261>
22. Kamburoğlu K, İlker Cebeci AR, Gröndahl HG. Effectiveness of limited cone-beam computed tomography in the detection of horizontal root fracture. *Dent Traumatol* 2009; **25**: 256–61. doi: <https://doi.org/10.1111/j.1600-9657.2009.00770.x>
23. Avsever H, Gunduz K, Orhan K, Uzun İsmail, Özmen B, Egrioglu E, et al. Comparison of intraoral radiography and cone-beam computed tomography for the detection of horizontal root fractures: an in vitro study. *Clin Oral Investig* 2014; **18**: 285–92. doi: <https://doi.org/10.1007/s00784-013-0940-4>
24. Wenzel A, Haiter-Neto F, Frydenberg M, Kirkevang L-L. Variable-resolution cone-beam computerized tomography with enhancement filtration compared with intraoral photostimulable phosphor radiography in detection of transverse root fractures in an in vitro model. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology* 2009; **108**: 939–45. doi: <https://doi.org/10.1016/j.tripleo.2009.07.041>
25. Shelley AM, Brunton P, Horner K. Subjective image quality assessment of cross sectional imaging methods for the symphyseal region of the mandible prior to dental implant placement. *J Dent* 2011; **39**: 764–70. doi: <https://doi.org/10.1016/j.jdent.2011.08.008>
26. Agency HP. *Guidance on the safe use of dental cone beam CT (computed tomography) equipment*. HPA-CRCE-010. Chilton: Health Protection Agency; 2010. <https://www.gov.uk/government/publications/dental-cone-beam-computed-tomography-safe-usage>.
27. Altman DG. *Practical statistics for medical research*. London: Chapman and Hall/CRC press; 1991.
28. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44**: 837–45. doi: <https://doi.org/10.2307/2531595>
29. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983; **148**: 839–43. doi: <https://doi.org/10.1148/radiology.148.3.6878708>
30. Lee CS, Nagy PG, Weaver SJ, Newman-Toker DE. Cognitive and system factors contributing to diagnostic errors in radiology. *American Journal of Roentgenology* 2013; **201**: 611–7. doi: <https://doi.org/10.2214/AJR.12.10375>
31. Waite S, Scott J, Gale B, Fuchs T, Kolla S, Reede D. Interpretive error in radiology. *American Journal of Roentgenology* 2017; **208**: 739–49. doi: <https://doi.org/10.2214/AJR.16.16963>
32. Gunderman RB. Biases in radiologic Reasoning. *American Journal of Roentgenology* 2009; **192**: 561–4. doi: <https://doi.org/10.2214/AJR.08.1220>
33. Rohlin M, Horner K, Lindh C, Wenzel A. Through the quality kaleidoscope: reflections on research in dentomaxillofacial imaging. *Dentomaxillofac Radiol* 2020; **19**: 20190484. doi: <https://doi.org/10.1259/dmfr.20190484>
34. Melo SLS, Bortoluzzi EA, Abreu Jr. M, Corrêa LR, Corrêa M. Diagnostic ability of a cone-beam computed tomography scan to assess longitudinal root fractures in prosthetically treated teeth. *J Endod* 2010; **36**: 1879–82. doi: <https://doi.org/10.1016/j.joen.2010.08.025>
35. European Commission. *Radiation Protection 172. Evidence based guidelines on cone beam CT for dental and maxillofacial radiology*. Luxembourg: Office for Official Publications of the European Communities; 2012. <https://ec.europa.eu/energy/sites/ener/files/documents/172.pdf>.
36. Nemtoi A, Czink C, Haba D, Gahleitner A. Cone beam CT: a current overview of devices. *Dentomaxillofac Radiol* 2013; **42**: 20120443. doi: <https://doi.org/10.1259/dmfr.20120443>
37. Talwar S, Utneja S, Nawal RR, Kaushik A, Srivastava D, Oberoy SS. Role of Cone-beam Computed Tomography in Diagnosis of Vertical Root Fractures: A Systematic Review and Meta-analysis. *J Endod* 2016; **42**: 12–24. doi: <https://doi.org/10.1016/j.joen.2015.09.012>