MRI Detection of Vertical Root Fractures in Endodontically Treated Teeth

A THESIS

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY

Beth R. Groenke, DDS

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

Donald Nixdorf, DDS, MS

Laurence Gaalaas, DDS, MS

Alex Fok, PhD

August 2020

Acknowledgements

A very special thank you to Dr. Donald Nixdorf, my primary advisor. You have been there for me every step of the way throughout this project and have always believed in me, pushed me, and given me the freedom to accomplish more than I thought possible.

Thank you to my committee members, Drs. Alex Fok and Laurence Gaalaas. You both provided excellent guidance throughout the research process.

Dr. Djaudat Idiyatullin, thank you for your patience in helping me learn the nuances of MRI and for your dedication to finding the best parameters possible.

Dr. Ashley Petersen, thank you very much for the statistical analysis insights and for your quick turn-around times. It was much appreciated.

Dr. Estephan Moana-Filho was an invaluable resource through various stages of the project, and I appreciate the suggestions and advice along the way. You helped tremendously at critical times. Thank you.

To my raters, Drs. Alan Law, Brian Barsness, and Mathew Royal, thank you for being available and for your generous, donated time.

To the UMN TMD and Orofacial Pain Family, I love you all. There is no way I could have accomplished this without your support through the program.

Thank you to the staff, faculty, and students on the 16th floor who were always there to help with questions, especially: Dr. Hooi Pin Chew, Dr. Young Heo, Bonita Vanheel, and Dr. Wondwosen Aregawi.

Thank you to Drs. Scott McClanahan and Ronald Ordinola Zapata for preliminary advice and equipment.

A special thank you to my friends and co-residents who have been there with me through the day-to-day highs and lows of clinical training and research: especially Connor Peck, Asha Sude, Mayank Shrivastava, Aaron Nisley, Anas Al-Taee, Barbara Botorous, and Anne Beaudin and the rest of the periodontics, prosthodontics, and endodontics residents. I wish I had space to name you all.

Thank you to Joseph Morrey for the encouragement to pursue this program and for your support leading to this point.

Finally, I would like to thank my parents, David and Sharon Groenke, for their unconditional love and support, not only during my training, but throughout my life. They have always encouraged me to pursue my dreams to the fullest extent.

Abstract

INTRODUCTION: Vertical root fracture (VRF) is known to occur in root canal treated (RCT) teeth and results in tooth loss. VRFs are difficult to diagnose. Magnetic Resonance Imaging (MRI) has the potential to identify VRF due to beneficial partial volume averaging, without using ionizing radiation. This investigation aims to compare the sensitivity and specificity of MRI versus cone-beam computed tomography (CBCT) in detecting VRF, using micro-computed tomography (microCT) as the reference standard. It also will describe the limits of MRI for detecting VRF.

METHODS: 115 extracted human tooth roots were RCT using common techniques. VRFs were induced in a proportion that resulted in 62 VRF samples and 53 non-fractured control samples. All samples were imaged in a phantom using MRI and CBCT. Axial images for MRI and CBCT were presented to three board-certified endodontists. Evaluators determined VRF status and a confidence assessment for that decision. 30% of images were resampled to calculate intra- and inter-rater reliability. For MRI, the most coronal slice with discernible VRF was measured on correlated microCT to determine the minimum VRF width (µm).

RESULTS: Sensitivity for MRI and CBCT were 0.66 (95%CI:0.53-0.78) and 0.58 (95%CI:0.45-0.70). Specificity was 0.72 (95%CI:0.58-0.83) and 0.87

(95%CI:0.75-0.95). Intra-rater reliability ranged from k=0.29-0.48 for MRI and k=0.30-0.44 for CBCT. Inter-rater reliability for MRI was k=0.37 and CBCT k=0.49. Median VRF width detected using MRI was 39µm (first quartile:20µm, third quartile: 58µm).

CONCLUSION: MRI demonstrated ability to repeatedly detect VRF as small as 20 µm. There was no significant difference between sensitivity nor specificity for MRI versus CBCT in detecting VRF, despite the early stage of MRI development.

Keywords: Dentistry, Tooth, Diagnosis, SWIFT, Pain

GRANT SUPPORT:

American Academy of Orofacial Pain

NIH: P41-EB027061, S10-RR023730

Research reported in this publication was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health Award Number UL1-TR002494. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Table of Contents

List of Tables

List of Figures

List of Equations

INTRODUCTION

Vertical root fractures (VRF) are longitudinal discontinuities in tooth roots that run parallel to the long axis of the tooth and extend from the apex toward the dental crown (1). The incidence rate is unknown. However, VRF are found in clinical practice with a prevalence ranging from 3.7% to 31% in root canal treated (RCT) teeth (2, 3). The most recent 2006 American Dental Association Survey on Dental Services Rendered reported over 15.1 million root canals were completed annually in the United States, indicating VRFs are not rare (4). Furthermore, VRF is specifically responsible for 6.4% to 25.3% of tooth extractions following RCT. There is no apparent gender predilection; and, VRFs are reported more frequently in patients in middle to upper ages, meaning 40 years old and older (5-8). As the United States population continues to age and people are retaining their teeth longer, the impact of VRF on society is likely to increase.

If left untreated, the probability for inflammation, infection, destruction of the supporting periodontium, and development of centralized pain increases (2, 9, 10). Complicating the diagnosis of VRF, signs and symptoms are not pathognomonic, meaning the typical dental evaluation cannot reliability detect VRFs. Common odontogenic disorders for which VRF are often mistaken include endodontic pathosis and periodontal lesions (1). Common confounding diagnoses of non-odontogenic origin can include: painful temporomandibular disorders, as well as neuropathic, neurovascular, or neoplastic pains. Present treatment for a confirmed VRF is typically extraction due to poor long-term prognosis; in contrast, management for patients with non-odontogenic pain

typically requires a multidisciplinary team approach with use of systemic medications. Furthermore, surgical interventions in this vulnerable patient population may be especially harmful and contribute to worsening of symptoms (9-11). Because treatment modalities for odontogenic versus non-odontogenic pain vary so drastically, and are associated with different prognoses, obtaining a proper diagnosis is essential to direct treatment to the source of the problem.

The most reliable diagnosis of VRF is made by direct visualization of the tooth root following surgical exploration or by radiographic identification of the fracture itself with accompanying late changes in surrounding bone (2, 12, 13). However, neither of these modalities are ideal. When the VRF location is positioned palatally, lingually, or intraradicullarly, the fracture may not be accessible and/or visible to the surgeon, thus preventing a definitive diagnosis (14). Furthermore, surgical exploration is costly, resource intensive, time and skill dependent, and an invasive procedure that carries associated risks for the patient. Alternatively, one method of indirect VRF identification is cone beam computed tomography (CBCT), which has been highlighted as becoming the most accurate imaging method for diagnosis of VRF versus conventional two-dimensional radiography (15-17). However, the diagnostic capabilities of CBCT are also limited. One limitation is that VRF often occurs in teeth with RCT and/or with dental posts present, and these filling materials generate substantial artifact within the CBCT image; thus complicating the diagnosis under the precise circumstances in which deriving the diagnosis of a VRF is most critical (18-21). Additionally, VRF diagnosis using CBCT relies heavily on identification of

associated bone loss from the surrounding structures (13, 22-24). Therefore, an indirect imaging method of detecting VRF early within its developmental process, meaning while it is small in dimensions, is needed. If VRFs are detected early, surgical intervention at such a time would allow for retention of the tooth itself or a decision could be confidently made to extract the tooth earlier, thus preserving the periodontium and facilitating future restorative treatment in addition to preventing patient suffering (12, 13, 24).

Several indirect modalities have been explored for their capability to detect VRF in addition to CBCT including: conventional dental radiography, optical coherence tomography (OCT), and magnetic resonance imaging (MRI) (25). Conventional radiography has been compared to CBCT in multiple studies, and while 2D radiography has demonstrated higher specificity than CBCT in some studies, it has been shown to have very low sensitivity and is considered unreliable to detect VRF (21, 23, 26).

OCT utilizes the principle of differences in optical reflectivity generated from infrared light waves to form images. It has demonstrated satisfactory sensitivity and specificity for detecting VRF with high resolution (~1-15µm) in real time; however, depth of penetration into the sample is limited to 2 mm, and the fracture must be accessible to the light source for visualization (27). While OCT has been used intra-operatively to diagnose VRF during RCT from within the canal space, it is not able to clinically evaluate apical tooth structure *in vivo* once the canal has been sealed. Therefore, OCT's use is limited to confirm a suspected VRF prior to RCT completion, to *ex vivo* studies, or to instances where

the fracture has already propagated coronally to the extent it is visible intraorally or via surgical exploration (28-30).

MRI is one method that shows promise for detecting VRF (31, 32). MR imaging has been used in the medical field since the 1980's (33). Recently, exploration has begun expanding regarding uses for MRI within the dental field (31, 32, 34-40). Given that VRF are filled with fluid and/or soft tissues with many loosely packed protons, and because the adjacent dental structures have fewer and more tightly packed protons, it is hypothesized that the signal contrast between the VRF and adjacent tooth structure will be greater than that for CBCT, thus facilitating identification of small VRF (25, 32). Furthermore, due to the effects of partial volume averaging, it was shown in a pilot study that MRI was able to detect VRF as small as 20 µm in width (32, 41). However, at this time the sensitivity and specificity for MRI detection of VRF is unknown. A more robust estimate of detectible fracture size is also needed.

Fryback and Thornbury (1991) presented a 5-step model to determine the efficacy for new diagnostic imaging methods that is being followed with this research. The steps they outlined and basic concepts each level seeks to answer were:

level 1-technical efficacy: What can be physically seen with the modality?

level 2-diagnostic accuracy efficacy: How well is the image interpreted?

level 3-diagnostic thinking efficacy: How helpful is the modality clinically?

level 4-therapuetic efficacy: How helpful is the modality to plan care?

level 5-patient outcome efficacy: Are patient outcomes improved by using

the modality?

As dental MRI is still in the developmental stages, this study aims to address level 1 and level 2 evidence that includes evaluating the sensitivity and specificity of MRI to detect VRF as well as determining the area under the curve (AUC) (42).

The first goal of our present study was to compare the sensitivity and specificity of MRI for detecting VRF to that of CBCT, using microCT as the reference standard for both imaging modalities. The second aim was to describe the limits of detection for MRI, specifically VRF width in microns, based on the opinion of 3 board-certified endodontists, in extracted endodontically treated teeth with microCT serving as a reference standard. Our hypothesis was that MRI would be more sensitive and specific than CBCT in detecting VRF in extracted, endodontically treated root samples. We also hypothesized that MRI would have a limit of detect VRF no narrower than 10-20 µm.

METHODS

Institutional Review Board at the University of Minnesota determined this study to be exempt from oversight because it was deemed not involving human subjects by DHHS and FDA regulations.

A team of researchers collaborated throughout the study and will be referenced via abbreviated text: Beth Morrey, DDS (BM), Djaudat Idiyatullin, PhD (DI), Laurence Gaalaas, DDS, MS (LG), Ashley Petersen, PhD (AP), Alan Law, DDS, PhD (AL), Brian Barsness, DDS, MS (BB), Mathew Royal, DDS (MR), Ronald Ordinola-Zapata, DDS, PhD (RZ), Alex Fok, PhD (AF), Wondwosen Aregawi, PhD (WA), Donald R. Nixdorf, DDS, MS (DN).

Sample Size Determination

Sample size was determined based on preliminary studies from our group that suggested MRI had a VRF detection sensitivity of 0.80. We generated a table of 95% confidence interval (CI) widths, with the minimum number of samples necessary to achieve said CI widths, to estimate the number of fractured samples required to describe VRF dimensions at those various levels. Based on this information, and for feasibility purposes, we selected a size of 60 VRF samples (Table 1).

Confidence Level	Sample Size (N)	Target Width	Actual Width	Proportion (P)	Lower Limit	Upper Limit	Width if P = 0.5
0.950	381	0.100	0.100	0.500	0.450	0.550	0.100
0.950	366	0.100	0.100	0.600	0.549	0.649	0.102
0.950	320	0.100	0.100	0.700	0.648	0.748	0.109
0.950	245	0.100	0.100	0.800	0.745	0.845	0.124
0.950	141	0.100	0.100	0.900	0.839	0.939	0.163
0.950	93	0.200	0.199	0.500	0.400	0.600	0.199
0.950	89	0.200	0.199	0.600	0.496	0.696	0.203
0.950	78	0.200	0.199	0.700	0.591	0.790	0.217
0.950	60	0.200	0.200	0.800	0.682	0.882	0.245
0.950	37	0.200	0.199	0.900	0.763	0.962	0.307
0.950	39	0.300	0.299	0.500	0.350	0.650	0.299
0.950	38	0.300	0.297	0.600	0.442	0.740	0.303
0.950	33	0.300	0.299	0.700	0.530	0.829	0.323
0.950	26	0.300	0.297	0.800	0.613	0.910	0.359
0.950	17	0.300	0.297	0.900	0.678	0.975	0.429

Table 1: Number of samples required for 95% confidence level assuming 80% sensitivity of MRI for detecting VRF

Additionally, 40 non-fractured samples were included as controls for sensitivity and specificity analysis to obtain a total of 100 samples. Unequal distribution of teeth was chosen to limit the total number of samples to be imaged and interpreted, thereby reducing cost and rater burden. We prepared an additional 20 samples to provide a margin of error to insure sufficient numbers of samples present in the final sample set because some amount of loss was expected (*i.e*. shattered samples, imaging complications, *etc*.).

Sample Preparation

Intact, permanent human teeth that were extracted for non-study related purposes were collected within the School of Dentistry and stored in 10% buffered formalin, starting with the collection process and maintained wet in this solution throughout the entire study. Exclusion criteria included teeth with prior RCT, third molars, and teeth with obvious anatomical abnormalities, such as

longitudinally fused roots, bulbous roots, or sharply curved roots. We decoronated the teeth at the cementoenamel junction and resected 2 mm of tooth apices to remove the apical arborization of the root. This removed most distinguishing tooth characteristics that could influence rater responses and allowed for standardization of the fracturing and imaging processes.

The same operator (BM) performed root canal treatment on each root segment using the recommended clinical protocol for WaveOne Gold (Dentsply Sirona). To maintain clinical relevance, we irrigated with 6% sodium hypochlorite (Chlorox) throughout the procedure and performed a final 17% EDTA (Vista Dental Products) rinse prior to obturation. We obturated the canals using Pulp Canal Sealer (Kerr) and GuttaCore (Dentsply Sirona). We removed extruded apical gutta percha with a high-speed rotary carbide bur. Following RCT we inspected the samples via 2.5x magnification and optical coherence tomography (Santec IVS-2000 F80, Nenjozaka, Japan) to identify if preliminary VRF developed during the RCT procedure. If no VRF was suspected the tooth was assigned to the control group. If VRF was suspected, the sample was assigned to the fracture group.

Samples in the fracture group were individually mounted in Imprint Heavy Body Vinyl Polysiloxane (3M ESPE) with the apices directed upward. As every root had unique density, form, and size, therefore widely varying fracture mechanics, we individually induced VRF using an 858 MTS Mini Bionix II (MTS Systems Corporation, Eden Prairie, MN) by manually applying progressively increasing force to the apex via a tapered 0.25 mm bunted probe inserted into

the apical foramina. We advanced the probe 0.5 mm and let it to settle in 5 to 10 second increments to permit microscopic deformation to occur, thus allowing stable fracture propagation without cleavage fracture (Figure 1). We monitored the sample for fracture initiation throughout this process using transillumination. When signs of VRF were observed (audible or visual), we immediately lifted the probe to terminate application of the load, thereby generating small VRF more akin to that which would be expected to form in early VRF. We manually punctured control samples at the apical foramina using the same probe to similarly distort the gutta percha at the apex to minimize the risk of rater bias.

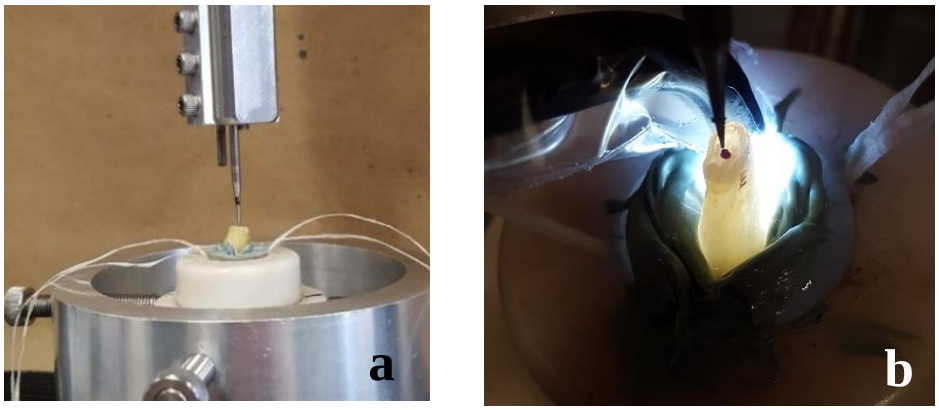

Figure 1: Fracture induction setup: a) apparatus setup, b) transillumination with probe

We generated a variety of fractures ranging from very small to completely cleaved. If cleaved samples could be easily reapproximated, we manually reduced the VRFs and secured the sample together by wrapping the fragments with wet Kimwipes (Kimberly-Clark Professional, Neenhah, Wisconsin) prior to placing them in vials. All samples were kept in individual glass vials (diameter: 21

mm; length: 70 mm) (Acme Vial & Glass Co., Inc., Paso Robles, CA, USA) and maintained in a single position by packing around the sample with Kimwipes.

Imaging of Samples

We imaged all samples with three different methods, all at the University of Minnesota. MicroCT reference imaging, to determine true VRF dimensions, was obtained using a Metrix 3D microCT (model XT H 225; Nikon Metrology, Brighton, MI), housed within the Minnesota Dental Research Center for Biomaterials and Biomechanics (MDRCBB) within the School of Dentistry. Imaging parameters were: 80 kV, 95 µA, for 708 ms with 720 projections and 4 frames/projection for each sample. MRI imaging was obtained using a 4.0-Tesla (T) whole-body MRI scanner (90 cm bore, Oxford, UK) with an Agilent DirectDrive console (Palo Alto, CA), housed at the Center for Magnetic Resonance Research (CMRR). The pulse sequence used was Sweep Imaging with Fourier Transformation (SWIFT) with the following parameters: scan time = 5 m 39 s, field-of-view (FOV) = 120 mm x 120 mm x 120 mm, flip angle = 12.0°, relaxation time = 5.21 ms, bandwidth = 407 Hz, oversampling = 512, pulse length = 4.8 µs, delay after pulse = 4.5 µs, power = 59 dB, voxel size = 0.2 mm isotropic. CBCT imaging was obtained with a clinical small FOV Carestream scanner (Carestream–CS 9000, Rochester, NY), housed within the Clinical Systems of the School of Dentistry. Imaging parameters were: 70 kV, 8.0 mA, 10.8 s, voxel size = 0.076 mm. For comparability reasons, both the MRI and

CBCT images were acquired using a purpose-built phantom that served to hold the samples (8 samples total per scan) and mimic the human head (Figure 2).

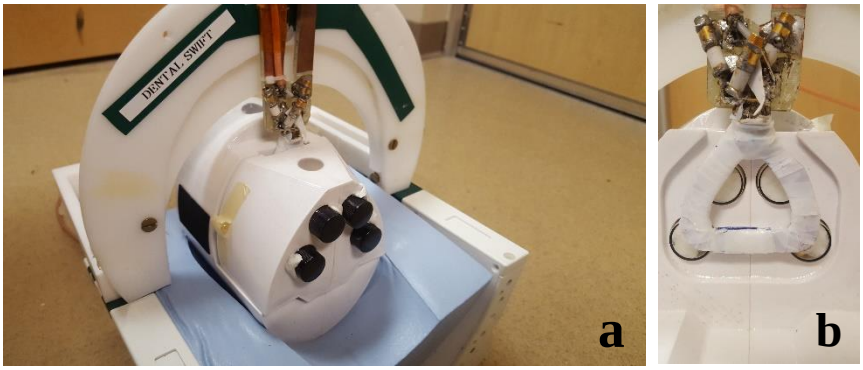

Figure 2: Study phantom a) entire apparatus, b) mandibular portion removed to allow visualization of alignment of intraoral coil

Image Preparation

We adjusted both MRI and CBCT images to visually appear to raters to be of similar size, which was approximately twice the size of the actual root. We cropped all images to include only the glass vial and sample for both MRI and CBCT to further standardize viewing.

Due to a lack of commercial imaging software, initial reconstructed MRI images exhibited off-resonance blurriness that varied in position throughout each sample. Using ImageJ™ (software version 1.52t, National Institutes of Health, USA), we constructed focused image stacks for all MRI samples using the glass edge of the vial as a reference to concatenate an uninterrupted stack of on-resonance images, similar to that which would be done automatically in clinical imaging software programs.

We used the Windows-based RadiAnt DICOM (64-bit) Viewer (software version 5.5.1.23267, Medixant, Poznan, Poland) to evaluate and crop CBCT images to the region of interest.

We obtained microCT images based on optimum achievable resolution for each specimen based on its size; therefore, the number of axial image slices varied per specimen. To properly correlate with same number of MRI axial slices, we individually scaled microCT axial stacks in ImageJ™ using bilinear interpolation by changing the “Depth (images)” parameter to match the number of slices in the corresponding MRI sample using VRF geometry and root morphology as guides for alignment.

Measurement of VRF Width

We measured VRF width using the scaled microCT stacks in ImageJ™ at the slice number provided by the endodontist evaluators based on MRI. For fractures extending the entire diameter of the specimen, 2 independent representative areas were measured; one on each side of the canal, and averaged. For fractures extending from the canal through half of the sample, one representative area was analyzed (Figure 3).

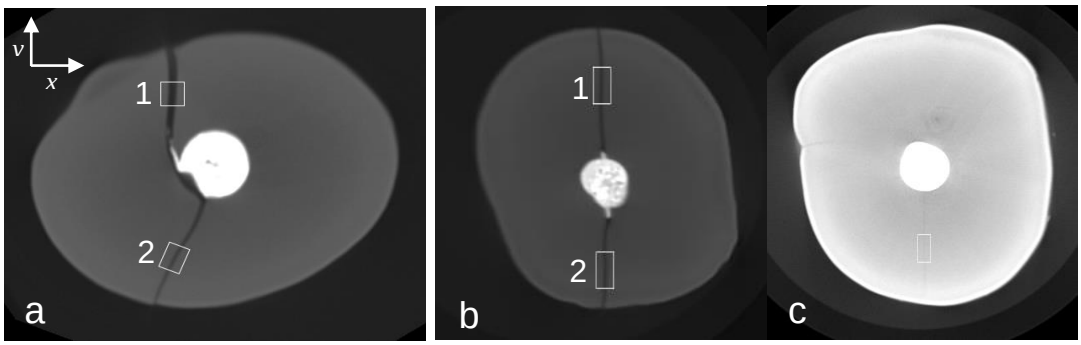

Figure 3:Examples of measured areas. a) oblique sample showing full fracture and within sample VRF variation (1. 51 µm, 2. 17 µm, mean 40 µm, b) vertical sample showing uniform full fracture at rater selected slice (1.22.5 µm, 2. 22.8 µm, mean 23 µm), c) same vertical sample as b showing half-fracture at smallest scaled microCT slice (3 µm). Contrast adjusted to optimize VRF appearance.

The most direct way to measure the VRF width at the selected slice would be to measure the width of the VRF's rectangular gray scale profile oriented orthogonally to the VRF. However, such measurement is highly inaccurate if the width of VRF is comparable or less than the width of the point spread function (PSF). This is because any imaging modality is characterized by a PSF that describes the response of an imaging system to a point source or point object. The image of any object is a convolution of the ideal image of the object with the PSF function. The width of the VRF's are comparable with the width of PSF; therefore, the VRF's profile in this case does not have an expected rectangular shape, but rather is a rectangular function smoothed by a low pass filtering that appears more similar to a Gaussian function. The width obtained by direct measurements in this case is not accurate and will be an overestimation of the real width of the VRF. However, an integral of the area underneath a VRF profile is not affected by low-pass filtering and will always be equal to the integral over dentin lowered by the value equal to dentin signal missing the VRF area. Based

on this property, the method using integrals described below was used to calculate the width of VRF.

To calculate the VRF width (w), we oriented the scaled microCT slices that correlated with the rater-selected MRI slices to align the VRF longitudinally with the conventional y-axis (Figures 3, 4). We used the “rectangle” feature in ImageJ™ to select a representative area of the VRF to analyze. We insured that the selected area encompassed the entire VRF profile (f) by checking the "plot profile" feature, then recorded the exact width (m) of this selection area using the "list" feature within plot profile. We also recorded the mean density of the selected area using ImageJ™’s measurement feature. Three additional means were obtained using the same selection box: two immediately adjacent (right (R) and left (L)) of the VRF to obtain mean density of variations in dentin signal surrounding the fracture, and one of the background noise (b). We then calculated fracture width using Equation 1 below.

$$VRFw = \left[1 - \left(\frac{\left(\int f - \int b\right)}{\left(\left(\frac{\left(\int R + \int L\right)}{2} \right) - \int b \right)} \right) \right] \cdot m$$

Equation 1: Method to calculate VRF width

In other words, by maintaining a consistent rectangular area for all measurements, and after correcting for background noise, a ratio of the overall contribution of signal from dentin plus VRF to dentin alone was determined. Because the VRF width was intentionally correlated to align with the x-axis, the proportion of signal loss attributed to VRF could be extrapolated to determine VRF width by multiplying by the known total width of the measured rectangular box (Figure 4).

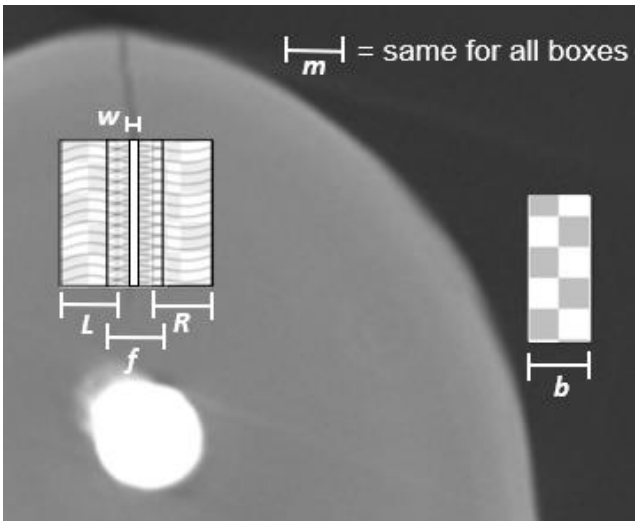

Figure 4: Visual representation of integral areas

To evaluate the error introduced by scaling, both full stack and scaled stack axial microCT images, were evaluated independently by a board-certified radiologist (LG) and a board-certified orofacial pain specialist (DN), both familiar with the technology. For the same specimen in both the full and scaled stacks, the evaluators determined the most coronal slice with the smallest detectible VRF, or in instances of distorted, oblique fractures, determined the most representative slice comparable to the most coronal slice of the scaled stack. In

instances of disagreement, evaluators used consensus to choose one slice. We measured microCT VRF widths at the consensus slices for both the full and scaled stacks and recorded the absolute value of the difference between these values and measures of spread, as well as the correlation coefficient. This measurement allowed us to estimate scaling distortion between the full and scaled microCT image stacks.

To account for variation in sample positioning during image acquisition and for potential image misalignment with MRI compared to scaled microCT slices, additional VRF measurements were obtained immediately adjacent to the identified VRF (*i.e*. once slice above and one slice below). We calculated the absolute value of change in VRF width measured from the additional slices for a subset of 22 samples. We selected a range of 3 slices, approximately 0.5 mm total, since many small VRF were only evident in a total of 2 to 3 slices of the sample. This measurement allowed us to estimate apical inaccuracies due to sample position.

Rater Training and Calibration

We chose raters who are board certified endodontists, clinically active, and familiar with three-dimensional imaging. Trainers were a board certified oral maxillofacial radiologist, an orofacial pain specialist, and an orofacial pain graduate student.

We held training and calibration sessions where axial MRI and CBCT image stacks were provided to the raters to familiarize them with the imaging

modalities and to standardize criteria for identification of VRF to achieve a goal of 80% agreement for intra-rater reliability and inter-rater reliability. See Appendices A and B for detailed training and calibration instructions as well as VRF identification criteria.

All training, and subsequent rating, sessions were held in the Oral Radiology interpretation clinic within the School of Dentistry using standardized diagnostic workstations with Dell 19” P1913S LCD monitors (screen resolution 1280 x 1024) and subdued ambient lighting. We confirmed monitors performance prior to each rating session using the AAPM TG18-QC test image to confirm visibility of both the 5% and 95% contrast levels, with all monitors passing each time. Raters used ImageJ™ (v1.52a) to view and manipulate all images.

Each session contained a set series of images for the raters to evaluate. MRI and CBCT images were reviewed in alternating stacks of 10 samples per modality to limit rater drift. The daily starting set of images (MRI or CBCT) was determined at random. All raters reviewed all images. Raters were allowed to scroll through axial image stacks but were not allowed to adjust contrast settings.

We asked raters to determine if a discontinuity, that being a VRF, was present in each MRI and CBCT specimen using a binary response (Yes/No). We also requested raters provide a confidence rating for that assessment using a 4-point Likert scale where: 1 was “not confident,” 2 was “somewhat confident,” 3 was “confident,” and 4 was “very confident.” We used this information to generate a receiver operant characteristic (ROC) curve of fracture detection based on rater confidence. For MRI, if a fracture was reported, the rater recorded the most

apical and most coronal slices where the VRF discernible and made a hand-drawn likeness of the fracture to record the shape and orientation of the fractures. We used this to more thoroughly discriminate Type I, or false positive, errors. In cases of disagreement between the 3 raters, a 2 of 3 majority was used to determine the response.

Statistical Analyses

A statistician (AP) performed statistical analysis using R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria). Using microCT as a reference standard, we performed sensitivity and specificity calculations for the MRI and CBCT data sets as well as subset sensitivity and specificity analyses between sessions and raters to assess for drift over time and/or rater differences. We used McNemar's test to explore for significance between differences in MRI and CBCT. Area under the curve (AUC) was calculated from the ROC curve generated by plotting sensitivity and 1-specificity for all possible thresholds applied to the consensus confidence assessment rating level to obtain binary classifications (predicted VRF verses not fractured). We calculated DeLong's test for correlated ROC curves. To assess raters' use of the confidence scale, we plotted number of teeth with VRF verses all teeth with that confidence rating to evaluate for trends.

Descriptive statistics (median, interquartile range, minimum and maximum values) were utilized as appropriate, based on non-parametric data distribution. To err on the side of providing more conservative measurements and to account

for errors introduced during scaling and stack alignment, we calculated total error by combining the absolute error between the full and scaled microCT stacks and the average absolute change in VRF width.

For MRI and CBCT, we determined intra-rater reliability using Cohen's kappa based on 30% of resampled images. 100% of the sample was used to determine inter-rater reliability using Cohen's kappa to explore individual rater pairings. Reliability among all 3 raters was assessed using Fleiss' kappa. We set agreement levels at: <0.00 poor, 0.00-0.20 slight, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 substantial, 0.81-1.00 excellent (43).

RESULTS

Samples Processed

We prepared 120 roots with decoronation, root tip amputation, and RCT. 72 samples had VRF induced, which resulted in 3 samples experiencing catastrophic failure, meaning the sample fractured into more than 2 pieces and they had to be discarded because they could not be reapproximated. We obtained a final sample of 69 roots with VRF and 48 roots without VRF serving as control samples. Additional controls allowed for the possibility of discovering very subtle VRF, following microCT imaging, anticipating that a small subset of VRF occurring from sample preparation would be undetectable until that stage. After microCT imaging, we discovered 6 samples that underwent VRF induction, did not meet criteria for VRF based on microCT criteria. Furthermore, we found two samples with residual, obvious anatomical characteristics following imaging and these samples were subsequently used only for training/calibration purposes. This resulted in a final sample number of 115, with 62 having verified VRFs and 53 controls.

Imaging Analysis and MRI/microCT Correlations

MicroCT image stacks contained an average of 1,356 slices/sample (range: 839-1,818 slices) while MRI image stacks averaged 58 slices/sample (range: 41-80 slices). Using the bilinear interpolation method, we scaled full axial microCT image stacks to individually correlate with the same number of slices in MRI image stacks for that sample.

Two primary types of VRF based on orientation were identified in the study during the scaling process: primarily non-oblique (n=53) and primarily oblique (n=9) VRF. Of note, 6 of the 9 oblique VRFs were samples that were complete VRF, meaning VRFs that resulted in 2 pieces which were reapproximated. The oblique VRFs exhibited more distortion after scaling (mean width change ± SD: 4.1 ± 19 µm smaller for scaled stack) than those in the non-oblique group (mean width change ± SD: 0.6 ± 3 µm larger for scaled stack); however, this was not statistically significant; therefore, we pooled all data ($p = 0.48$).

The measurement error introduced by scaling the pooled full microCT axial image stacks to the scaled microCT axial image stacks resulted in scaled microCT stack widths that were 0.6 µm smaller on average compared to the full microCT stack widths (minimum: -33 µm, interquartile range: (-2) µm - 1.75 µm, maximum: 36 µm) We did not find this scaling difference to be statistically significant ($p = 0.71$). Furthermore, we found good correlation between full and scaled microCT stacks (correlation = 0.89) and especially good correlation for non-oblique VRFs (correlation = 0.96) (Figure 5).

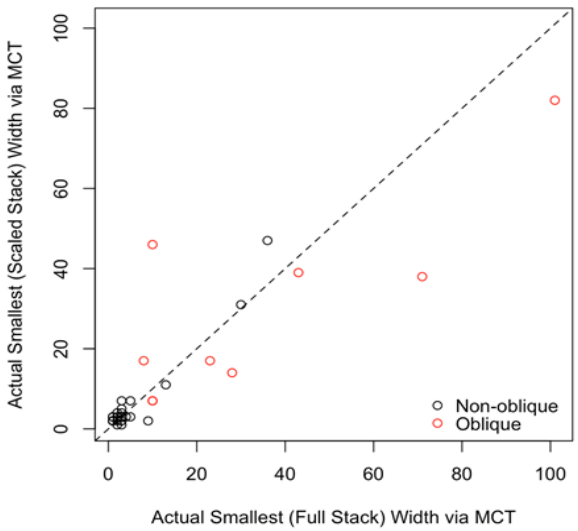

Figure 5: Plot of full vs.scaled stack microCT.. Red dots denote oblique fractures which were more distorted than non-oblique fractures, but this was not statistically significant

To account for variation in physical positioning of the samples between imaging modalities and to account for discrepancies introduced from aligning the scaled mircoCT and MRI image stacks, we also summarized the average absolute change in width over 3 consecutive scaled microCT slices using the average rater identified slice as the middle reference point. Median change in VRF width between 3 slices was 2.5 µm (minimum change: 0.5 µm, interquartile range: 2.0 µm - 3.4 µm, maximum: 34 µm). While VRFs typically propagate in a "V" pattern, narrowing toward the coronal aspect of the sample, we found that the course is not always purely linear. Therefore, small variations in VRF width may occur, leading to a transient, counter-intuitive widening of the VRF as it is measured coronally.

We estimated total measurement error by combining the absolute error between the full and scaled microCT image stacks and the average absolute

change in width for 3 consecutive microCT slices. Median total error was 5.3 µm (minimum: 1.5 µm, interquartile range: 3.0 µm - 14.3 µm, maximum: 41.5 µm).

Rater Performance

Three of the authors were selected to perform evaluation of the images (AL, BB, MR) due to their status as board-certified endodontists and familiarity with 3D imaging techniques in their clinical practice. They have a combined clinical experience of 60 years.

A total of 3 calibration sessions were held to achieve the desired 80% inter-rater agreement threshold. Ratings were obtained over 5 rating sessions spanning a maximum of 27 days (Rater 1 (R1): 9 days, Rater 2 (R2): 27 days, Rater 3 (R3): 14 days), held at the evaluators' convenience, and lasting no more than 2 hours per day. Sixty images were reviewed each session: 30 MRI and 30 CBCT.

Intra-rater Reliability

Intra-rater reliability ranged from fair to moderate for both MRI and CBCT: k = 0.29 to 0.48 and k = 0.30 to 0.44 respectively and overall agreement remained consistent when applying restricted criteria and adjusting for orientation-based false positives (Table 2).

	MRI	CBCT	Restricted MRI
Rater 1	0.29	0.44	0.27
Rater 2	0.42	0.30	0.37
Rater 3	0.48	0.31	0.54

Table 2: Intra-rater reliability

Inter-rater Reliability

Inter-rater reliability using Fleiss's kappa to evaluate the relationship between all raters for MRI was fair (k = 0.37) and moderate for CBCT (k = 0.49). Cohen's kappa was used to sub-analyze the reliability between individual rater pairs. There was similar inter-rater reliability between all pairs of raters indicating no single rater over-influenced the general results. Inter-rater reliability improved overall to moderate-substantial (k = 0.55 to 0.67) when adjusted for restricted MRI criteria, thus eliminating orientation based false positives (Table 3). As raters were not requested to specify orientation of CBCT images, no direct comparison could be made.

	MRI	CBCT	Restricted MRI
Rater 1 vs. 2	0.41	0.54	0.55
Rater 2 vs. 3	0.60	0.55	0.66
Rater 1 vs. 3	0.61	0.55	0.67
All raters	0.37	0.49	0.49

Table 3: Inter-rater reliability

Sensitivity and Specificity

For the first aim, comparing the sensitivity and specificity of MRI to CBCT using microCT as the reference standard for both modalities, sensitivity for MRI and CBCT were 0.66 (95% CI: 0.53 to 0.78) and 0.58 (95% CI: 0.45 to 0.70) respectively. When more restrictive criteria were applied for MRI by eliminating false-positive responses based on misidentified VRF, per rater VRF sketches, sensitivity for MRI was 0.55 (95% CI: 0.42 to 0.68). Direct comparison to CBCT sensitivity based on restrictive criteria was not possible as raters were not asked

to sketch final CBCT slices due to time constraints during rating sessions (Table 4).

Specificity for MRI was 0.72 (95% CI: 0.58 to 0.83) with and without the restrictive criteria as and 0.87 (95% CI: 0.75 to 0.95) for CBCT (Table 4). We confirmed that there was no evidence of a statistically significant difference between MRI and CBCT in terms of sensitivity (p = 0.27) or specificity (p = 0.12) using McNemar's test.

	MRI	CBCT	Restricted MRI
Sensitivity	**0.66** (0.53-0.78)	**0.58** (0.45-0.70)	**0.55** (0.42-0.68)
Specificity	**0.72** (0.58-0.83)	**0.87** (0.75-0.95)	**0.72** (0.58-0.83)

Table 4: Sensitivity and specificity of MRI, CBCT, and restricted MRI (Est (95% CI))

Subset sensitivity and specificity analyses between all five rating sessions and between all three raters were evaluated. There were no significant trends to indicate drift over time or rater differences significantly impacted the outcomes. Details of this analysis can be found in Appendices C - F.

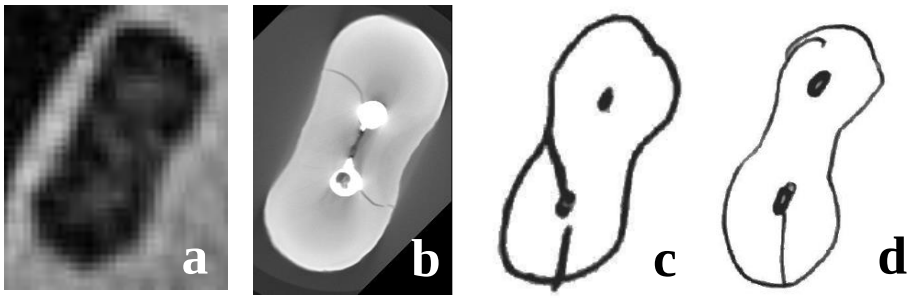

Figure 6: Misidentified sample based on restricted MRI criteria. a) MRI sample, b) microCT, c & d) rater drawings showing fracture in location other than that shown on microCT

Confidence Scoring

We evaluated the raters' overall use of the confidence assessment scale. Table 5: Consensus confidence rating of VRF identification, shows the number of teeth with VRF out of the total number of teeth with that confidence rating along with the percentage. No consensus ratings ever "very confidently" stated there was no VRF and overall confidence was lower for ruling out VRF for both CBCT and MRI. Conversely, when raters reported they were "very confident" there was a VRF, the were correct 100% of the time for CBCT and correct 80% of the time for MRI. When raters were less confident in their decisions for both MRI and CBCT. Figure 7: Confidence Assessment Scale Use shows the same data in graphical format.

	No VRF				VRF			
	Very Confident	Confident	Somewhat Confident	Not Confident	Not Confident	Somewhat Confident	Confident	Very Confident
MRI	0/0	1/10 (10%)	13/29 (45%)	7/21 (33%)	9/15 (60%)	9/11 (82%)	11/14 (79%)	12/15 (80%)
CBCT	0/0	2/10 (20%)	12/33 (36%)	13/26 (50%)	5/13 (38%)	4/6 (67%)	11/12 (92%)	15/15 (100%)

Table 5: Consensus confidence rating of VRF identification: total teeth with VRF in each category/total teeth with that rating (ratio expressed as percentage)

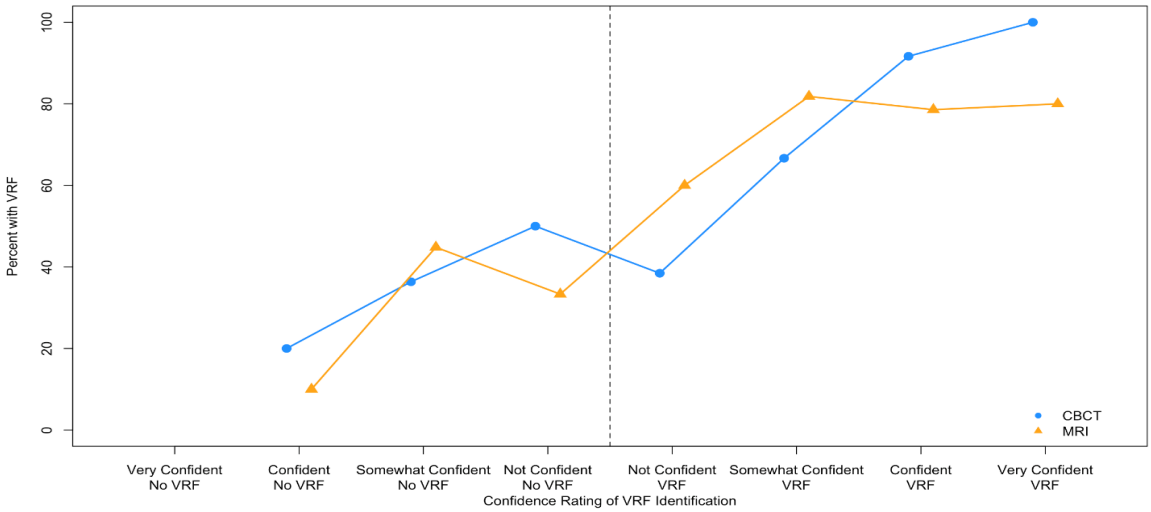

Figure 7: Consensus confidence rating of VRF identification

ROC Curve

A receiver operant curve (ROC) of fracture detection based on rater confidence was generated comparing CBCT and MRI. The area under the curve (AUC) was 0.75 (95% CI: 0.66 to 0.84) for CBCT and 0.74 (95% CI: 0.65 to 0.83) for MRI. Using DeLong's test for correlated ROC curves, there was no evidence of a statistically significant difference between the AUCs for CBCT and MRI (p = 0.88) (Figure 8).

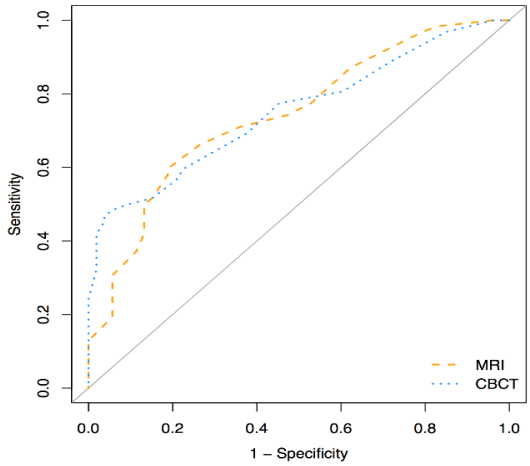

Figure 8: ROC of MRI vs CBCT

Limits of VRF Width Detection

Of the 53 fractured samples, consensus raters correctly identified 34 of them as having VRF using MRI. The VRF widths measured from the scaled stack microCT that corresponded to the last slice the raters could visualize the VRF on the MRI had a median width of 39 µm (minimum: 3 µm, interquartile range: 20 µm - 58 µm, maximum: 93 µm) (Figure: 9). Taking into account the median calculated total error of 5 µm based on sample misalignment, the smallest detectible VRF widths could feasibly fall within 15 µm – 63 µm. Figure 10 shows examples of various fracture sizes detected by MRI with a representative CBCT slice and the corresponding microCT slice.

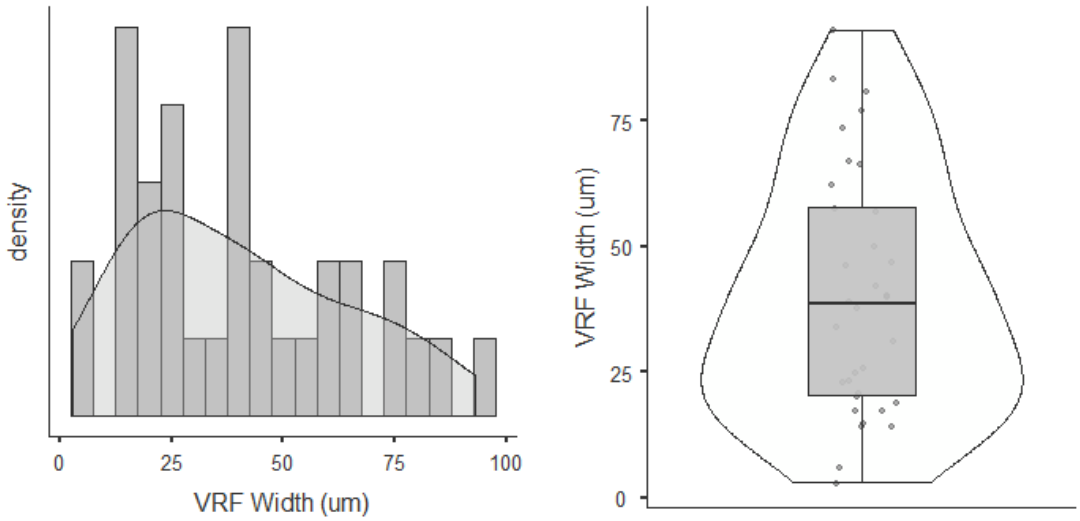

Figure 9: VRF distribution of identified VRF width by rater consensus (µm)

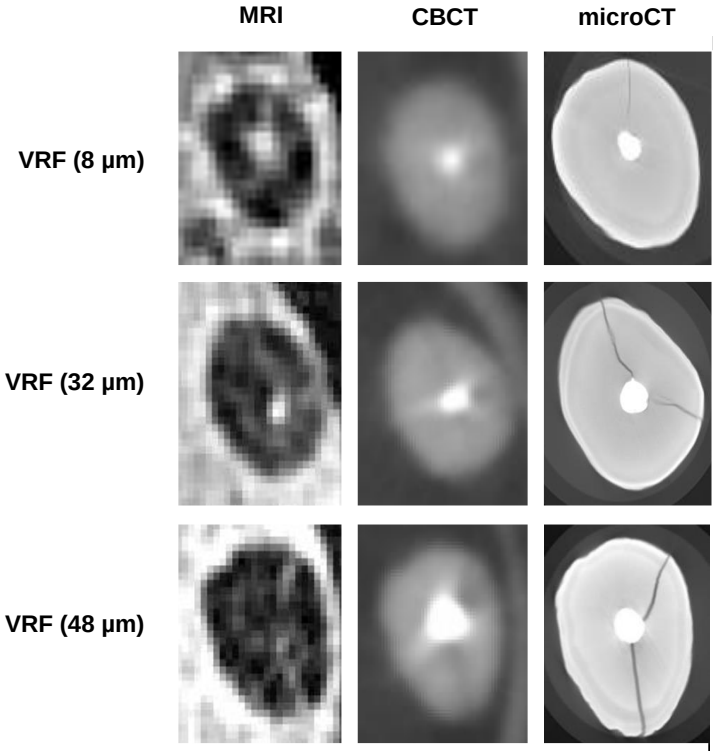

Figure 10: Example of imaging modalities and varying VRF sizes

DISCUSSION

Our study found no significant difference with measures of sensitivity or specificity for MRI verses CBCT in detecting VRF in extracted endodontically treated sample teeth roots when microCT served as the reference standard. These findings are similar to previous observations from our group that used teeth suspected of having cracks and VRF and extracted during the course of clinical care and inspection was used as the reference standard (25). Evaluators demonstrated ability to detect VRF on MRIs that had a median width of 39 µm and interquartile range being 20 to 58 µm in size. As expected, the smallest VRF width detected by raters was larger than the calculated theoretical limit of detection for MRI, that being 10 to 20 µm (32).

Intra- and Inter-Rater Reliability

The only other comparable study examining MRI reliability to the authors' knowledge was the previous work by our group examining both clinically occurring VRF and cracked teeth who reported fair-to-excellent intra-rater reliability (k = 0.38 to 1.0) for MRI and moderate-to-excellent (k = 0.66 to 1.0) for CBCT (25). Inter-rater reliability in the same study was reported to be fair (0.21) for MRI and moderate (0.45) for CBCT, which is the comparable to our present findings. Inter-rater reliability improved for both MRI and CBCT from our last study, which may be due to rater training and familiarity with the system, since 2 of our raters were also part of the previous study. Intra-rater reliability; however, was lower in the present study and could be attributed to data collection sessions

being held over several days-to-weeks, as opposed to one day as was done previously, and the task being to detect small VRF. Overall, inter-rater agreement is similar to our previous report while intra-rater reliability was reduced for MRI and CBCT.

Additional *ex vivo* studies have explored intra- and inter-rater agreement for CBCT in which VRF were artificially induced using an Instron machine, but from a crown-downward approach, as opposed to the apices upward approach used in our methods. Two of those studies placed a gutta percha (GP) cone within the canal. Both studies classified VRFs as either complete or incomplete. The intra-rater agreement ranged from 0.27 to 0.55 and from 0.70 to 0.87 and inter-rater agreement 0.005 and 0.35 to 0.51 respectively (21, 44). Intra-rater agreement with the first study was similar to our findings; however, the intra-rater agreement was higher in the second comparison study versus our findings. The first study used 3 endodontists and 3 endodontic residents as raters while the second used 5 radiologists. Our study used 3 board-certified endodontists. Differences in training and experience between raters may have accounted for the distinct differences in inter-rater agreement, especially considering the low inter-rater reliability cited in the first group. A third study used the same method to induce VRF as the previous 2 studies but did not place GP. They evaluated two different CBCT systems' ability to diagnose VRF (26). Their study also categorized fractures as both complete or incomplete. Intra-rater reliability for complete fractures for the two systems was 0.80 and 0.78 and for incomplete VRF 0.50 and 0.67. Inter-rater reliability was 0.60 and 0.59 and 0.23 and 0.33 for

complete and incomplete VRF respectively. The decreased intra-rater reliability for CBCT in the present study compared to other studies, with the exception of one that was comparable, was likely due to the presence of smaller, therefore harder to detect, VRFs. Additionally, use of dental pulp canal sealer in our study, as opposed to only the placement of a GP cone in other studies, may have generated more artifact that enhanced difficulty with detecting VRF.

Additional *ex vivo* studies have been performed exploring intra- and inter-rater agreement with regard to diagnostic capability of CBCT to detect VRF; however, those studies are inhomogeneous in sample composition (*i.e.* fiber posts, metal posts, gutta percha, *etc*.) and report reliability measures as a single statistic across their heterogenous samples (19, 20, 45, 46). Therefore, individual results will not be described as they may be misleading in the context of our uniform sample.

Confidence Scoring

Analysis of our confidence assessment scale showed that raters used the entire spectrum of confidence levels while performing their examinations. To assist in this occurring, we intentionally did not provide a neutral category to help raters avoid the expected behavioral response of giving a neutral response under uncertain conditions. In general, despite having a higher actual percentage of VRF samples than controls, 52% of MRI responses and 61% of CBCT consensus responses indicated a level of confidence that there was no VRF. However, when raters gave a “very confident” rating for CBCT images they were

100% correct with their assessment, and for MRI images they were 80% correct in detecting a VRF being present. This means when rater confidence was high regarding the presence of a VRF, they were almost always correct. VRF with a “very confident” rating were often larger in dimension and/or oblique in character as opposed to those VRF with lower confidence ratings. For VRF that the raters were “not confident” about, raters tended to be less confident in general that there was no VRF and used assessment level “not confident” and “somewhat confident” most frequently. Additionally, there was no consensus score for “very confident no VRF” assigned when they did not identify a VRF for either CBCT or MRI. We attribute this to the difficult clinical task that is identification of VRF, especially when exploring the small fractures generated in this study. Overall, use of the confidence assessment scale showed that raters were consistent in using the scale itself for both MRI and CBCT and were more confident while identifying larger and oblique VRF.

Sensitivity and Specificity

The only available study, to the authors’ knowledge, exploring sensitivity and specificity of MRI to detect VRF is our prior study that showed an MRI sensitivity of 0.59 versus the MRI sensitivity of 0.66 reported here. MRI specificity was 0.83 in the prior iteration opposed to 0.72 in the present study. Our current study optimized MRI image resolution to the highest extent presently achievable; however, this resulted in more image noise, observed as graininess in the image stacks. The increased sensitivity and decreased specificity versus the prior study

could be attributed to false-positive responses by the raters from this noise being mistaken for very small VRF. We applied more restricted MRI VRF identification criteria by verifying that the VRFs identified by the raters were correct via comparison of VRF drawings to the microCT gold standard. With this standard applied, we showed an MRI sensitivity of 0.55, which was more similar to that reported previously. For CBCT, the prior study also reported a sensitivity of 0.59 while this study showed a very similar sensitivity of 0.58. Previous CBCT specificity was 0.90 compared to the 0.87 reported here. Differences in this study could be attributed to the fact that our prior study explored teeth extracted due to suspicion of naturally occurring fractures. It is possible that fractures which are already symptomatic are larger, and therefore easier to detect, than the small, artificial fractures simulated in this study. Furthermore, the prior study had a mix of RCT and non-RCT samples and the presence of intra-canal materials may have affected the ability of raters to detect VRF under those differing conditions (20, 44, 45).

Overall, present sensitivity and specificity results for MRI and CBCT were similar to those previously reported.

Four previous projects have explored the *ex vivo* sensitivity and specificity of CBCT for detecting VRF in the presence of gutta percha (GP) with induced VRFs. VRFs in 2 studies were induced with a hammer and chisel and VRF was determined based on presence under a stereomicroscope (19, 45). They reported CBCT sensitivities of 0.79 and 0.94 and specificity of 0.88 and 1.00 respectively. A third study also induced VRF using hammer and chisel, but used

transillumination to verify VRF induction (46). They explored the effect of 2 CBCT voxel sizes and different intracanal materials for the ability of CBCT to diagnose VRF. In their sub-group of samples most similar to those in our study, using voxel size of 0.2 mm and presence of GP, they reported sensitivity and specificity of 0.93 and 0.73 respectively. Due to VRF induction method and gold standard utilized, VRFs in these studies may have been different, likely larger, than those in our study which would explain the higher overall sensitivity and specificity. The other study exploring CBCT diagnosis of VRF with GP used an Instron machine, but induced VRF from a crown-down approach versus an apex upward approach as used in our study (21). They sub-categorized fractures as complete and incomplete and reported sensitivity of 0.69 and 0.53 respectively. The incomplete VRF results in that study closely match the CBCT sensitivity of 0.58 in our study. In contrast, they reported specificity of 0.37 for both complete and incomplete fractures, while our study showed a specificity of 0.87. This difference may be attributed to the use of a human mandible in their study: trabecular patterns of bone could be mistaken for VRF and lead to lower specificity. Additionally, they used PowerPoint to show raters images, which has less functionality than if raters can directly manipulate images, as was allowed in our study. Of note, our study was the only *ex vivo* study to use root canal sealer. Presence of a radiopaque sealer, which would be present clinically, may cause more distortion and lower sensitivity and specificity values in our study compared to others. Despite this, sensitivity and specificity for both MRI and CBCT in our present

study falls within the range of values reported previously in the literature for similar studies.

ROC Curve

Area under the curve has been calculated for three similar *ex vivo* studies exploring VRF detection using CBCT; however, comparison studies were not available for MRI. All of these studies induced the VRF using an Instron machine from the crown-downward approach. One CBCT study placed GP in 20 samples with induced VRF that ranged from 50 to 110µm wide based on OCT. They categorized their results based on complete and incomplete VRF. Interestingly, they reported an AUC of 0.45 for both complete and incomplete VRF, which was lower than our observation of 0.75 for CBCT (21). Another study from that same group using similar methods, but without a GP cone, explored complete, incomplete, and VRF <50 µm and ≥50 µm using 2 CBCT systems. For complete fractures, AUC was 0.99 for both imaging systems and for incomplete fractures 0.69 and 0.66. For VRF ≥50 µm they reported AUCs of 0.78 for both systems and for VRV <50 µm 0.62 and 0.57 (26). Both of these studies utilized PowerPoint to show raters study images versus a traditional imaging viewer. The second study from that group did not include GP, so the enhanced visibility from lack of image artifact may have offset any hinderance from the PowerPoint format, leading to their improved accuracy. Additionally, both of those studies used endodontic residents as well as fully trained endodontists as raters, which may affect accuracy. A third study prepared 30 samples with RCT and used

transillumination to distinguish between complete and incomplete VRF. They explored the effects of both intra-canal posts and GP on CBCT diagnosis using different CBCT imaging settings and sub-divided VRFs into complete and incomplete. For the subset of VRF+GP samples pertinent to our study, they reported AUC values ranging from 0.47 to 0.57 for incompletely separated VRF samples and from 0.77 to 0.88 for completely separated samples, depending on CBCT imaging mode (44). Expectedly, in that study, images obtained using higher resolution modes had greater accuracy than images taken using standard imaging parameters. Although direct comparison data was not available for MRI, because MRI and CBCT AUC were not significantly different in our study, and because our CBCT AUC data is similar or better than that previously published, it illustrates that even though MRI is a developing technology, its diagnostic capabilities already fall within the higher range of imaging systems that are being used clinically for identifying small VRFs.

Limits of VRF Width Detection

We found 4 studies that previously reported specific VRF widths, and all pertained to CBCT. One study used microCT as a reference standard; however, they reported VRF data on length only and therefore details will not be discussed (17). Of the studies that reported VRF width, one study used electronic calipers to set known fracture widths to be <200 µm (separated root fragments glued back together) or≥200 µm or ≥400 µm using calipers and wax to measure and maintain the segments at the designated widths. CBCT was determined to be

effective at detecting VRFs of different thicknesses (47). This is not surprising as fracture widths in this study were quite large, and is expected that CBCT would be able to detect these differences (48). Three studies used OCT to measure VRF width at the widest portion of the fracture, contrary to our study that sought to describe the narrowest portion of the fracture detectible. Our study found the narrowest median VRF width identified by MRI to be 39 µm. One study using OCT reported VRF widths ranging from 60 to 770 µm. They selected a cutoff value for fractures <300 µm vs ≥300 µm and found no significant difference in CBCT detection for fractures above or below this threshold (23), which is not surprising given the relatively large threshold value selected. Another study categorized simulated VRF as either complete or incomplete. They reported incomplete VRF widths ranged from 50 to 110 µm and determined that CBCT was not reliable in detecting induced VRF (21). In a subsequent study, OCT was used to measure the widest portion of identified VRF and widths were reported that ranged from 30 to 110 µm. Sensitivity/specificity analyses were divided for fractures <50 µm and ≥50 µm, and it was reported that detection of VRF by CBCT was significantly lower for fractures <50 µm for the 2 different CBCT systems analyzed (26). Our results were measured based on mircoCT as a reference and examined the smallest portion of the fracture so are not directly comparable; however, this prior literature supports that CBCT is limited in its ability to detect very subtle VRF. Furthermore, while OCT is an efficient method to screen and/or measure VRF with high resolution in real time, its diagnostic

capabilities are limited to portions of a tooth that are visible to its light source, thus preventing OCTs use for *in vivo* VRF detection.

Study Limitations

A physical limitation of our study is that teeth needed to be imaged in a separate container for microCT, compared to CBCT and MRI. This resulted in positioning discrepancies between the reference standard images obtained from microCT and test imaging modalities. Additionally, stack alignment between scaled microCT and MRI stacks was performed using visual inspection based on VRF geometry and anatomical tooth outline. Therefore, taken together, some misalignment between the samples occurred. Efforts were made and reported to quantify this error; however, to increase the precision of future studies, it is recommended a calibration marker of known dimension be applied to each sample to facilitate alignment.

This experiment included several study improvements designed to more closely simulate reality; however, there are still advancements to be made. The purpose of this experiment was to push the boundaries of MRI for detecting VRF; to do so, we decreased voxel size as much as possible to generate high-resolution images capable of detecting very small fractures (32). However, this resulted in less signal per voxel, which reduced our overall signal-to-noise ratio (SNR). Because signal is directly proportional to the square root of background noise, relatively minor changes in signal loss resulted in visually appreciable differences in image quality. This lower SNR was the reason for grainy images,

which were harder for the raters to interpret based on their feedback. We attempted to mitigate some of the effects of poor SNR through training and calibration; however, it may have been insufficient and could be a contributing factor to variation in overall study results and rater agreement.

This study iteration also incorporated the use of an agar-filled phantom to simulate soft tissue attenuation for MRI and CBCT; whereas, the previous study imaged samples directly in glass vials. Adding attenuation artifacts could affect the presentation of very small structures such as early VRF. As the technology progresses, further iterations should consider incorporating a phantom to mimic boney structures, which was performed in several of our comparison studies.

Image artifacts are known to occur with the presence of gutta percha in CBCT imaging and background noise complicates MRI VRF diagnosis, we expanded this study by requesting raters draw a sketch of what they considered to be the fracture which was later verified with microCT to further scrutinize for false-positives. As expected, by using more restricted criteria, MRI sensitivity was reduced to 0.55. Direct comparison to CBCT is recommended for the next iteration; however, due to time constraints during rating sessions, was unable to be completed for this study.

Studies exploring VRF have used different measures to determine VRF presence or absence and VRF dimensions including but not limited to: transillumination, light microscopy, methylene blue dye, microCT, and OCT. No studies were found comparing the reliability or precision of these methods to each other. Therefore, a consistently used reference standard method for the

identification and measurement for VRF should be developed to determine the best way to quantify VRF characteristics in future studies exploring development of imaging systems.

Topics of Future Research

Improvements with the MRI technology itself are still needed as the raters' primary feedback regarding the images were their grainy appearance. Refinement of the intraoral coil design to improve signal acquisition, thereby improving the signal-to-noise ratio without altering other scanning parameters, would be a step toward obtaining more clear images. An automated software package capable of performing post-processing functions and rendering standardized MRI images also needs to be developed to improve image presentation and consistency. Additionally, development of computer learning algorithms to detect VRF could enhance detection. Finally, future studies should continue to move toward a more clinically relevant model which includes transitioning the technology to a 1.5T and/or 3T systems and examining the effects of bone and soft tissues on fracture detection *in vivo*.

CONCLUSIONS

Our study found no significant difference between sensitivity or specificity in detecting VRF in extracted endodontically treated sample roots for both MRI and CBCT when microCT served as a reference standard for both. On the MRIs, raters demonstrated the ability to detect VRFs with median widths of 39 µm, with interquartile range of 20 to 58 µm. This suggests that MRI continues to show promise as one potential methodology that can further be developed for the purpose of *in vivo* VRF detection to improve endodontic care of patients.

BIBLIOGRAPHY

1. Endodontists AAo. Cracking the Cracked Tooth Code: Detection and Treatment of Various Longitudinal Tooth Fractures. Endodontics Colleagues for Excellence [Internet]. 2008. Available from: http://www.aae.org/specialty/wp-content/uploads/sites/2/2017/07/ecfesum08.pdf.
2. Liao WC, Tsai YL, Wang CY, Chang MC, Huang WL, Lin HJ, et al. Clinical and Radiographic Characteristics of Vertical Root Fractures in Endodontically and Nonendodontically Treated Teeth. Journal of endodontics. 2017;43(5):687-93.
3. Varshosaz M, Tavakoli MA, Mostafavi M, Baghban AA. Comparison of conventional radiography with cone beam computed tomography for detection of vertical root fractures: an in vitro study. J Oral Sci. 2010;52(4):593-7.
4. Association AD. 2005-2006 Survey of Dental Services Rendered. 2007.
5. Almasri M. Assessment of extracting molars and premolars after root canal treatment: A retrospective study. Saudi Dent J. 2019;31(4):487-91.
6. Toure B, Faye B, Kane AW, Lo CM, Niang B, Boucher Y. Analysis of reasons for extraction of endodontically treated teeth: a prospective study. Journal of endodontics. 2011;37(11):1512-5.
7. Olcay K, Ataoglu H, Belli S. Evaluation of Related Factors in the Failure of Endodontically Treated Teeth: A Cross-sectional Study. Journal of endodontics. 2018;44(1):38-45.
8. Maddalone M, Gagliani M, Citterio CL, Karanxha L, Pellegatta A, Del Fabbro M. Prevalence of vertical root fractures in teeth planned for apical surgery. A retrospective cohort study. Int Endod J. 2018;51(9):969-74.
9. Okeson JP. Bell's Oral and Facial Pain. 2nd ed. Hanover Park, IL: Quintessence Publishing, USA; 2014.
10. Hryvenko I. Long-term Outcome of Patients with Persistent Pain Following Root Canal Treatment: The National Dental Practice-Based Research Network.: University of Minnesota; 2018.
11. Sharav YaB, Rafael. Orofacial Pain and Headache. Hanover Park, IL: Quintessence Publishing, USA; 2015.
12. Cohen S, Blanco L, Berman L. Vertical root fractures: clinical and radiographic diagnosis. Journal of the American Dental Association (1939). 2003;134(4):434-41.
13. Walton RE. Vertical root fracture: Factors related to identification. Journal of the American Dental Association (1939). 2017;148(2):100-5.
14. Saberi E, Mollashahi NF, Movasagh Z, Moghaddam AA, Mohammadi A. Value of CBCT in vertical root fracture detection in endodontically-treated teeth. Minerva stomatologica. 2017;66(2):69-74.
15. Baageel TM, Allah EH, Bakalka GT, Jadu F, Yamany I, Jan AM, et al. Vertical root fracture: Biological effects and accuracy of diagnostic imaging methods. Journal of International Society of Preventive & Community Dentistry. 2016;6(Suppl 2):S93-s104.

16. Bernardes RA, de Moraes IG, Hungaro Duarte MA, Azevedo BC, de Azevedo JR, Bramante CM. Use of cone-beam volumetric tomography in the diagnosis of root fractures. Oral surgery, oral medicine, oral pathology, oral radiology, and endodontics. 2009;108(2):270-7.
17. Wang S, Xu Y, Shen Z, Wang L, Qiao F, Zhang X, et al. The Extent of the Crack on Artificial Simulation Models with CBCT and Periapical Radiography. PLoS One. 2017;12(1):e0169150.
18. Valizadeh S, Vasegh Z, Rezapanah S, Safi Y, Khaeazifard MJ. Effect of Object Position in Cone Beam Computed Tomography Field of View for Detection of Root Fractures in Teeth with Intra-Canal Posts. Iran J Radiol. 2015;12(4):e25272.
19. Moudi E, Haghanifar S, Madani Z, Alhavaz A, Bijani A, Bagheri M. Assessment of vertical root fracture using cone-beam computed tomography. Imaging science in dentistry. 2014;44(1):37-41.
20. 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. Journal of endodontics. 2010;36(1):126-9.
21. 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(12):1140-52.
22. Zhang L, Wang T, Cao Y, Wang C, Tan B, Tang X, et al. In Vivo Detection of Subtle Vertical Root Fracture in Endodontically Treated Teeth by Cone-beam Computed Tomography. Journal of endodontics. 2019;45(7):856-62.
23. Chavda R, Mannocci F, Andiappan M, Patel S. Comparing the in vivo diagnostic accuracy of digital periapical radiography with cone-beam computed tomography for the detection of vertical root fracture. Journal of endodontics. 2014;40(10):1524-9.
24. PradeepKumar AR, Shemesh H, Jothilatha S, Vijayabharathi R, Jayalakshmi S, Kishen A. Diagnosis of Vertical Root Fractures in Restored Endodontically Treated Teeth: A Time-dependent Retrospective Cohort Study. Journal of endodontics. 2016;42(8):1175-80.
25. Schuurmans TJ. Assessing the Accuracy and Reliability of Root Crack and Fracture Detection in Teeth Using Sweep Imaging with Fourier Transform (SWIFT) Magnetic Resonance Imaging (MRI): University of Minnesota; 2017.
26. Brady E, Mannocci F, Brown J, Wilson R, Patel S. A comparison of cone beam computed tomography and periapical radiography for the detection of vertical root fractures in nonendodontically treated teeth. Int Endod J. 2014;47(8):735-46.
27. Machoy M, Seeliger J, Szyszka-Sommerfeld L, Koprowski R, Gedrange T, Wozniak K. The Use of Optical Coherence Tomography in Dental Diagnostics: A State-of-the-Art Review. J Healthc Eng. 2017;2017:7560645.
28. Yoshioka T, Sakaue H, Ishimura H, Ebihara A, Suda H, Sumi Y. Detection of root surface fractures with swept-source optical coherence tomography (SS-OCT). Photomed Laser Surg. 2013;31(1):23-7.

29. Shemesh H, van Soest G, Wu MK, Wesselink PR. Diagnosis of vertical root fractures with optical coherence tomography. Journal of endodontics. 2008;34(6):739-42.
30. Lee SH, Lee JJ, Chung HJ, Park JT, Kim HJ. Dental optical coherence tomography: new potential diagnostic system for cracked-tooth syndrome. Surg Radiol Anat. 2016;38(1):49-54.
31. Schuurmans TJ, Nixdorf DR, Idiyatullin DS, Law AS, Barsness BD, Roach SH, et al. Accuracy and Reliability of Root Crack and Fracture Detection in Teeth Using Magnetic Resonance Imaging. Journal of endodontics. 2019;45(6):750-5 e2.
32. Idiyatullin D, Garwood M, Gaalaas L, Nixdorf DR. Role of MRI for detecting micro cracks in teeth. Dentomaxillofac Radiol. 2016;45(7):20160150.
33. Edelman RR. The history of MR imaging as seen through the pages of radiology. Radiology. 2014;273(2 Suppl):S181-200.
34. Idiyatullin D, Corum C, Moeller S, Prasad HS, Garwood M, Nixdorf DR. Dental magnetic resonance imaging: making the invisible visible. Journal of endodontics. 2011;37(6):745-52.
35. Idiyatullin D, Corum CA, Nixdorf DR, Garwood M. Intraoral approach for imaging teeth using the transverse B1 field components of an occlusally oriented loop coil. Magn Reson Med. 2014;72(1):160-5.
36. Demirturk Kocasarac H, Geha H, Gaalaas LR, Nixdorf DR. MRI for Dental Applications. Dental clinics of North America. 2018;62(3):467-80.
37. Ludwig U, Eisenbeiss AK, Scheifele C, Nelson K, Bock M, Hennig J, et al. Dental MRI using wireless intraoral coils. Sci Rep. 2016;6:23301.
38. Wanner L, Ludwig U, Hovener JB, Nelson K, Flugge T. Magnetic resonance imaging-a diagnostic tool for postoperative evaluation of dental implants: a case report. Oral surgery, oral medicine, oral pathology and oral radiology. 2018;125(4):e103-e7.
39. Assaf AT, Zrnc TA, Remus CC, Khokale A, Habermann CR, Schulze D, et al. Early detection of pulp necrosis and dental vitality after traumatic dental injuries in children and adolescents by 3-Tesla magnetic resonance imaging. J Craniomaxillofac Surg. 2015;43(7):1088-93.
40. Hilgenfeld T, Juerchott A, Deisenhofer UK, Weber D, Rues S, Rammelsberg P, et al. In vivo accuracy of tooth surface reconstruction based on CBCT and dental MRI-A clinical pilot study. Clin Oral Implants Res. 2019;30(9):920-7.
41. Molen AD. Considerations in the use of cone-beam computed tomography for buccal bone measurements. Am J Orthod Dentofacial Orthop. 2010;137(4 Suppl):S130-5.
42. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making. 1991;11(2):88-94.
43. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.
44. Neves FS, Freitas DQ, Campos PS, Ekestubbe A, Lofthag-Hansen S. Evaluation of cone-beam computed tomography in the diagnosis of vertical root

fractures: the influence of imaging modes and root canal materials. Journal of endodontics. 2014;40(10):1530-6.
45. Hassan B, Metska ME, Ozok AR, van der Stelt P, Wesselink PR. Detection of vertical root fractures in endodontically treated teeth by a cone beam computed tomography scan. Journal of endodontics. 2009;35(5):719-22.
46. Melo SL, Bortoluzzi EA, Abreu M, Jr., Correa LR, Correa M. Diagnostic ability of a cone-beam computed tomography scan to assess longitudinal root fractures in prosthetically treated teeth. Journal of endodontics. 2010;36(11):1879-82.
47. Ozer SY. Detection of vertical root fractures by using cone beam computed tomography with variable voxel sizes in an in vitro model. Journal of endodontics. 2011;37(1):75-9.
48. Ozer SY. Detection of vertical root fractures of different thicknesses in endodontically enlarged teeth by cone beam computed tomography versus digital radiography. Journal of endodontics. 2010;36(7):1245-9.

Appendix A: Training and Calibration Exercise Instructions

Training and Calibration Exercise Instructions

1. Please review the criteria for MRI and CBCT crack/fracture presentation documents prior to starting the image review (see attached documents in digital folder).
    - You are encouraged to reference these documents while reviewing the images.

2. Items to keep in mind for the study:
    - Interpret ALL the available slices in each DICOM image file – i.e. the whole visible image as you see it, then make the determination if the image appearance is consistent with crack / fracture.
    - ***For each image:***
        i. Note the initial slice # that allows visualization of the crack/fracture and the final slice # that allows visualization of the fracture. Draw a representation of each slice in a crude drawing of what you see on the screen
        ii. OPTIONAL Free Response: note any questions, concerns, artifact observed, etc. When possible, give the slice # as a reference.
    - ***Fatigue:***
        i. Observers will self-determine their level of fatigue throughout the study. If you feel fatigued, please take a break at any time. Do not continue when fatigued.
        ii. General Guidelines:
            1. *<u>Image review periods should not exceed 60 continuous minutes without a break.</u>*

    - ***Questions:***
        i. Assistance or clarification can be requested at any time.

Appendix B: Criteria for CBCT/MRI Crack/Fracture Presentation

Criteria for CBCT/MRI Crack/Fracture Presentation

A separation of the adjacent root segments on multiple contiguous image slices without the continuation of the bright or dark signal line into the adjacent tissue (or water as was present in this *ex vivo* study design). This signal line must be observed within the confines of tooth structure, delineated by:

a. *External tooth surface:* bounded by enamel, the external surface of dentin, or cemental tissue depending on the level of axial slice.
b. *Internal tooth surface:* external extent of the pulpal cavity

For purposes of this study, the physical discontinuity must have the following criteria to be given a designation of a crack/fracture.

1. Signal line(s) must extend from the external boundary of the tooth toward the pulpal cavity (or vice versa)* on contiguous image slices.^
2. The overall contour of the external tooth surface and pulpal cavity must be maintained.**
   The signal line(s) must be visible near the apex of the tooth.#

   Explanation of stated criteria:

   A. **These cracks are thought to be clinically significant.*
   B. *^The criteria of multiple image slices allow for detection of an angled crack.*
   C. ***To prevent gross root discontinuities from mistakenly being classified as a crack.*
   D. *#To prevent other defects or craze lines from being misclassified as a VRF.*

Furthermore, it is important to differentiate cracks/fractures from other commonly encountered pathologic changes in tooth morphology that may be observed, and to rule out artifact and possible false positive identification. The following entities should be differentiated from cracks/fractures:

- Multiple streak artifacts from root fillings that traverse the root and adjacent tissue.
- Aberrations in dental anatomy or morphology which may include
  - Accessory, lateral, or secondary pulp canal(s)
  - Canal ramifications
- Physiologic or pathologic processes which may include
  - Root Resorption (Inflammatory, Replacement, Surface, Cervical, External, Internal)
  - Caries

Appendix C: Supplemental Sensitivity and Specificity Analyses, Sub-Analyses by Session

Table 6: Sensitivity and specificity (with 95% CI) for each session for MRI

	Session 1	**Session 2**	**Session 3**	**Session 4**	**Session 5**
Sensitivity	82% (48-98%)	47% (21-73%)	54% (25-81%)	100% (66-100%)	64% (35-87%)
Specificity	56% (21-86%)	78% (40-97%)	80% (44-97%)	71% (42-92%)	73% (39-94%)

Table 7: Sensitivity and specificity (with 95% CI) for each session for CBCT

	Session 1	**Session 2**	**Session 3**	**Session 4**	**Session 5**
Sensitivity	44% (14-79%)	64% (35-87%)	60% (32-84%)	46% (19-75%)	73% (39-94%)
Specificity	70% (35-93%)	100% (63-100%)	80% (44-97%)	92% (64-100%)	92% (62-100%)

Table 8: Sensitivity and specificity (with 95% CI) for each session for Restricted MRI

	Session 1	**Session 2**	**Session 3**	**Session 4**	**Session 5**
Sensitivity	55% (23-83%)	47% (21-73%)	38% (14-68%)	89% (52-100%)	57% (29-82%)
Specificity	56% (21-86%)	78% (40-97%)	80% (44-97%)	71% (42-92%)	73% (39-94%)

Appendix D: Supplemental Sensitivity and Specificity Analyses, MRI Sub-Analyses by Rater

Table 9: Sensitivity and specificity (with 95% CI) for each session for MRI for Rater 1

	Session 1	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**All sessions**
Sensitivity	91% (59-100%)	67% (38-88%)	62% (32-86%)	100% (66-100%)	71% (42-92%)	76% (63-86%)
Specificity	44% (14-79%)	33% (7-70%)	70% (35-93%)	57% (29-82%)	45% (17-77%)	51% (37-65%)

Table 10: Sensitivity and specificity (with 95% CI) for each session for MRI for Rater 2

	Session 1	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**All sessions**
Sensitivity	73% (39-94%)	53% (27-79%)	62% (32-86%)	89% (52-100%)	57% (29-82%)	65% (51-76%)
Specificity	44% (14-79%)	67% (30-93%)	60% (26-88%)	64% (35-87%)	73% (39-94%)	62% (48-75%)

Table 11: Sensitivity and specificity (with 95% CI) for each session for MRI for Rater 3

	Session 1	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**All sessions**
Sensitivity	64% (31-89%)	40% (16-68%)	46% (19-75%)	89% (52-100%)	71% (42-92%)	60% (46-72%)
Specificity	78% (40-97%)	89% (52-100%)	90% (55-100%)	71% (42-92%)	82% (48-98%)	81% (68-91%)

Appendix E: Supplemental Sensitivity and Specificity Analyses, CBCT Sub-Analyses by Rater

Table 12: Sensitivity and specificity (with 95% CI) for each session for CBCT for Rater 1

	Session 1	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**All sessions**
Sensitivity	44% (14-79%)	64% (35-87%)	60% (32-84%)	46% (19-75%)	91% (59-100%)	61% (48-73%)
Specificity	70% (35-93%)	100% (63-100%)	80% (44-97%)	92% (64-100%)	92% (62-100%)	87% (75-95%)

Table 13: Sensitivity and specificity (with 95% CI) for each session for CBCT for Rater 2

	Session 1	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**All sessions**
Sensitivity	56% (21-86%)	64% (35-87%)	60% (32-84%)	46% (19-75%)	55% (23-83%)	56% (43-69%)
Specificity	70% (35-93%)	100% (63-100%)	90% (55-100%)	77% (46-95%)	75% (43-95%)	81% (68-91%)

Table 14: Sensitivity and specificity (with 95% CI) for each session for CBCT for Rater 3

	Session 1	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**All sessions**
Sensitivity	44% (14-79%)	71% (42-92%)	60% (32-84%)	54% (25-81%)	73% (39-94%)	61% (48-73%)
Specificity	60% (26-88%)	75% (35-97%)	80% (44-97%)	69% (39-91%)	67% (35-90%)	70% (56-82%)

Appendix F: Supplemental Sensitivity and Specificity Analyses, Restricted MRI Sub-Analyses by Rater

Table 15: Sensitivity and specificity (with 95% CI) for each session for Restricted MRI Rater 1

	Session 1	Session 2	Session 3	Session 4	Session 5	All sessions
Sensitivity	64% (31-89%)	60% (32-84%)	46% (19-75%)	100% (66-100%)	64% (35-87%)	65% (51-76%)
Specificity	44% (14-79%)	33% (7-70%)	70% (35-93%)	57% (29-82%)	45% (17-77%)	51% (37-65%)

Table 16: Sensitivity and specificity (with 95% CI) for each session for Restricted MRI Rater 2

	Session 1	Session 2	Session 3	Session 4	Session 5	All sessions
Sensitivity	45% (17-77%)	47% (21-73%)	46% (19-75%)	89% (52-100%)	50% (23-77%)	53% (40-66%)
Specificity	44% (14-79%)	67% (30-93%)	60% (26-88%)	64% (35-87%)	73% (39-94%)	62% (48-75%)

Table 17: Sensitivity and specificity (with 95% CI) for each session for Restricted MRI Rater 3

	Session 1	Session 2	Session 3	Session 4	Session 5	All sessions
Sensitivity	55% (23-83%)	40% (16-68%)	38% (14-68%)	78% (40-97%)	57% (29-82%)	52% (39-65%)
Specificity	78% (40-97%)	89% (52-100%)	90% (55-100%)	71% (42-92%)	82% (48-98%)	81% (68-91%)