Dentomaxillofacial Radiology (2021) 50, 20210175

birpublications.org/dmfr

RESEARCH ARTICLE

The challenge of applying digital image processing software on intraoral radiographs for osteoporosis risk assessment

[1]Joanna Gullberg, [2]Ayman Al-Okshi, [1]Dalia Homar Asan, [1]Anita Zainea, [3]Daniel Sundh, [3,4,5]Mattias Lorentzon and [1]Christina Lindh

[1]Faculty of Odontology, Malmö University, Malmö, Sweden; [2]Department of Oral Medicine and Radiology, Faculty of Dentistry, Sebha University, Sebha, Libya; [3]Sahlgrenska Osteoporosis Centre, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, University of Gothenburg, Gothenburg, Sweden; [4]Region Västra Götaland, Department of Geriatric Medicine, Sahlgrenska University Hospital, Mölndal, Sweden; [5]Mary MacKillop Institute for Health Research, Australian Catholic University, Melbourne, VIC, Australia

Objectives: The purpose of this study was to evaluate rater agreement and the accuracy of a semi-automated software and its fully automated tool for osteoporosis risk assessment in intraoral radiographs.
Methods: A total of 567 intraoral radiographs was selected retrospectively from women aged 75–80 years participating in a large population-based study (SUPERB) based in Gothenburg, Sweden. Five raters assessed participants' risk of osteoporosis in the intraoral radiographs using a semi-automated software. Assessments were repeated after 4 weeks on 121 radiographs (20%) randomly selected from the original 567. Radiographs were also assessed by the softwares' fully automated tool for analysis.
Results: Overall interrater agreement for the five raters was 0.37 (95% CI 0.32–0.41), and for the five raters with the fully automated tool included as 'sixth rater' the overall Kappa was 0.34 (0.30–0.38). Intrarater agreement varied from moderate to substantial according to the Landis and Koch interpretation scale. Diagnostic accuracy was calculated in relation to reference standard for osteoporosis diagnosis which is T-score values for spine, total hip and femoral neck and presented in form of sensitivities, specificities, predictive values, likelihood ratios and odds ratios. All raters' mean sensitivity, including the fully automated tool, was 40,4% (range 14,3%–57,6%). Corresponding values for specificity was 69,5% (range 59,7%–90,4%). The diagnostic odds ratios ranged between 1 and 2.7.
Conclusion: The low diagnostic odds ratio and agreement between raters in osteoporosis risk assessment using the software for analysis of the trabecular pattern in intraoral radiographs shows that more work needs to be done to optimise the automation of trabecular pattern analysis in intraoral radiographs.
Dentomaxillofacial Radiology (2021) **50**, 20210175. doi: 10.1259/dmfr.20210175

Cite this article as: Gullberg J, Al-Okshi A, Homar Asan D, Zainea A, Sundh D, Lorentzon M, et al. The challenge of applying digital image processing software on intraoral radiographs for osteoporosis risk assessment. *Dentomaxillofac Radiol* 2021; **50**: 20210175.

Keywords: Dental Digital Radiography; Computer-Assisted Image Analysis; Data Accuracy; Osteoporosis; Risk assessment

Introduction

In 1994, osteoporosis was defined by WHO as a systematic skeletal disease characterised by low bone mass and microstructural deterioration of bone tissue with a consequent increase in bone fragility and susceptibility to fracture.[1] Osteoporotic fractures are common and approximately 50% of females over 50 years of age will sustain a fragility fracture during their remaining

Correspondence to: Mrs Joanna Gullberg, E-mail: joanna.gullberg@mau.se
Received 12 April 2021; revised 08 June 2021; accepted 06 July 2021

lifetime.[2] In 2010, three and a half million osteoporotic fractures occurred in the EU at an annual cost of 37 billion Euro.[3] The number of fractures is expected to rise to four and a half million and cost 46.8 billion Euro by 2025.[3] In females, the increase is most prominent after the age of 70. In Sweden, the annual number of osteoporotic fractures is almost 70,000 and vertebral fractures are the most common.[4]

Osteoporosis is one risk factor for fracture and the diagnosis of osteoporosis mainly relies on the measurement of bone mineral density (BMD) by dual-energy X-ray absorptiometry (DXA).[5] The operational definition of osteoporosis is based on the T-score for BMD assessed by DXA at the femoral neck or spine and is defined as a value for BMD 2.5 SD or more below the young female adult mean.[6] A major challenge in managing osteoporosis is identifying affected individuals before the condition is established and fracture has occurred. Even though bone density measurement plays a vital role in examining individuals with multiple risk factors in order to predict their fracture risk, there is currently no scientific evidence to support the use of bone density measurement as a screening method in healthy, middle-aged individuals.[7]

It is known that a correlation exists between BMD of the jaws and other skeletal sites,[8–11] and studies have also shown that patients with osteoporosis have an altered trabecular pattern in the jaws compared with normal subjects.[12] Based on these conclusions, methods have been developed and applied on intraoral dental radiographs, with the goal of identifying individuals at risk of or with osteoporosis. Visual assessment, categorising the trabecular patter into dense homogeneous, heterogeneous, or sparse homogeneous in intraoral radiographs of the premolar region of both upper and lower jaw showed a potential to identify females at risk of having osteoporosis[13] as well as for prediction of skeletal fractures.[14] Most adults in Sweden frequently visit their dentist where radiographs are taken on a regular basis[15] for diagnostic purposes, and an opportunity exists to incorporate osteoporosis risk assessment into dental clinical practice using radiographs already taken in the dental setting.

Visual assessments have the drawback of being rater dependent which means they rely on human perception, and thus some degree of error is likely inevitable even with experienced raters.[16] Digital image processing software have been applied to overcome this issue in an attempt to automate the process of analysis of trabecular pattern in intraoral dental radiographs.[12,17–20] However, most software are based on a semi-automated approach where a rater applies the region of interest on a specific part of the image. As little as 0.10 cm^2 of interdental bone of the premolar area on intraoral dental radiographs has been proved to be large enough to enable prediction of femoral and spinal BMD and thus osteoporosis.[21]

Most studies are performed on radiographs obtained under controlled conditions within academic research projects and not on radiographs obtained during everyday clinical work in a primary dental setting.[13,18] Furthermore, as for all diagnostic methods, agreement and accuracy are important concepts and agreement studies using digital image analysis software are infrequently reported.[22] Even though an image processing algorithm is applied, these are often based on a semi-automated approach that does not entirely neglect the influence of its user. To the best of our knowledge, no study has evaluated rater agreement when using software for analysis of the trabecular bone pattern in intraoral dental radiographs. Therefore, this study aimed to investigate rater agreement and the diagnostic accuracy of one semi-automated software and its fully automated tool in osteoporosis risk assessment based on the analysis of trabecular pattern in intraoral dental radiographs obtained in a primary dental setting.

Methods and materials

This is a retrospective rater-based study on agreement and reliability when using semi-automated software for analyses of bone tissue in intraoral radiographs. It was conducted, analysed, and reported in accordance with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS).[23]

Subjects

A population-based, prospective study (Sahlgrenska University hospital Prospective Evaluation of Risk of Bone fractures—The SUPERB study) was based in Gothenburg, Sweden, between the years 2013 and 2016. A national population register was used to identify females aged 75–80 years living within the greater Gothenburg area. Those who were ambulant, able to follow instructions in Swedish and had at least one hip that could be evaluated for BMD were included in the study.[24,25] The study included 3028 elderly females. All study subjects gave their informed consent, and the study protocol was approved by the Ethical Review Board at the University of Gothenburg (Dnr T297-15/Ad 929–12).

Dual-energy X-ray absorptiometry

Areal (a) BMD of the hip and spine was measured by dual-energy X-ray absorptiometry (DXA) device, Hologic Discovery A (S/N 86491) (Waltham, MA) on most participants (n = 2995). The National Health and Nutrition Examination Survey (NHANES) III reference database for femoral neck and total hip in 20–29 year-old Caucasian females as well as the Hologic material for lumbar spine consisting of 30-year-old Caucasian American females were used to calculate the corresponding T-scores.[26] Due to machine failure, a small proportion of females (n = 33) was measured with

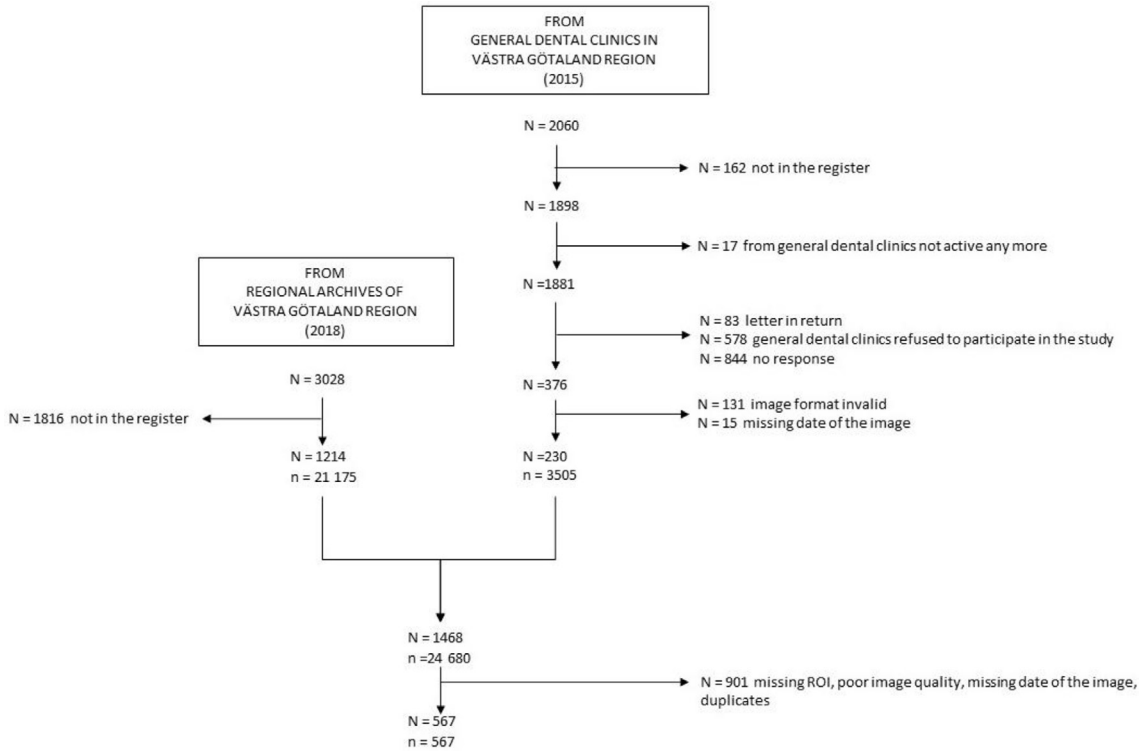

Figure 1 Flow chart of participant recruitment and data collection. *N*, number of study participants; *n*, number of dental radiographs; ROI, region of interest

another Hologic Discovery A DXA device (Waltham, MA). The potential discrepancy between the two machines was taken into account by performing a cross-calibration described elsewhere.[27]

Collection of dental radiographs

A flow chart illustrating the collection of images is shown in Figure 1. In 2015, data on 2060 participants out of 3028 from the SUBERB study[27] were available. The social security numbers of these study participants were sent to the Swedish National Insurance Agency requesting data on dental examinations during 2010–2015 in the Västra Götaland region of Sweden, including dental radiographic examinations. A total of 9303 dental examinations were found on 1898 participants from 337 clinics. Letters with lists of patients were sent to the clinics requesting that they send digital as well as analogue radiographs obtained as part of a dental examination. Radiographs from 376 patients were collected from 83 responding clinics. Invalid image format and missing date on the images reduced the amount to 230 study participants, 3505 images. Due to poor response rates from the clinics in the first round of collection of images in 2015, data on all 3028 participants from the SUPERB study were also requested from the regional archives of Västra Götaland that stores data for public healthcare clinics in 2018. We received images from 1214 patients and as we received all radiographic images from all examinations performed of the patient at the clinic, regardless of our inclusion criteria, this gave us a total of 21,175 images. Duplicates of images from the first and second data collection were removed. Thereafter, a selection of images took place according to the following criteria: (a) vertical bitewing and/or periapical image including ROI between roots of premolars in the lower jaw (b) acceptable image quality including projection geometry, resolution, sharpness and contrast (c) image taken within 3 years before or after DXA examination. First, a rough selection was performed by the first author. The most common reasons for excluding images from the analysis were missing region of interest, poor image quality, image too old, or unknown date. Two authors made the second selection of images based on image quality and closest in date to DXA examination. After the final selection 567 images, from the same number of patients remained for analysis. All DICOM images were converted to JPG format using Image J software (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, 1997–2018). Analogue images were scanned with 1000 dpi using UMAX Mirage IIse (Umax Technologies, Inc., Hsinchu, Taiwan) flatbed scanner.

Digital image analysis software

The software (Boneprox©, Gothenburg, Sweden) is based on a bilinear filter where the reference size is chosen to be small enough to remove noise but still retain the trabecular pattern to be assessed automatically. The

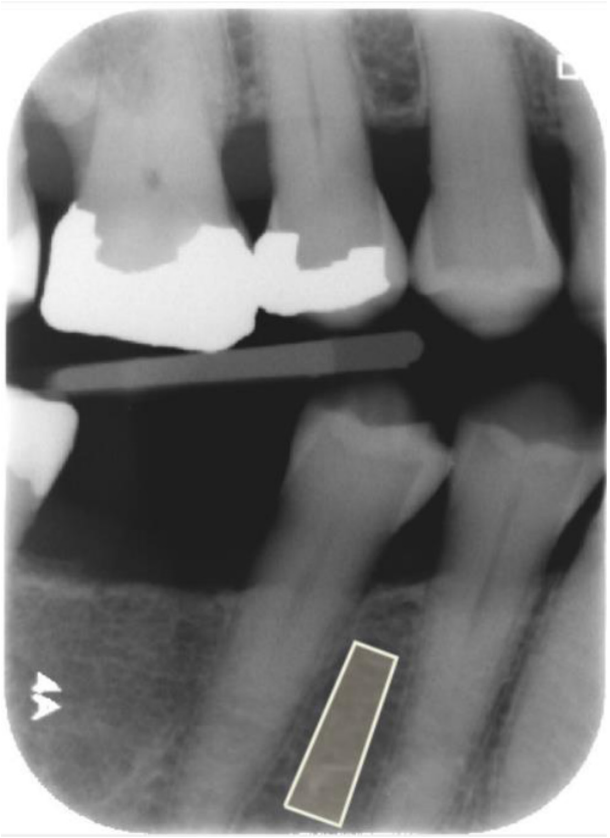

Figure 2 An example of a dental radiograph uploaded to the software with the trapezoid marker symbolising ROI manually placed between the roots/apices of premolar area in the lower jaw. ROI, region of interest.

grey levels in the image are reduced to 8-bit data and linearly stretched to cover all intensities. A median filter minimises noise in the image. The analysis progresses by identifying the minimum and mean value within a 9 × 9 neighbourhood to every pixel and focusing only on the bone tissue. The software searches for a pattern to recognise trabecula and separates them from the voids. If too many pixels are classified as a trabecula (and thus too little space) the threshold is reduced, and the process repeated. The threshold value was chosen from the reference material.[13] Afterwards, the distance transform is applied, first to measure from each pixel within every trabecula the closest bone space and then measure the size of the space between trabeculae. The largest intertrabecular space is identified, after that the second largest, until the 20 largest spaces are found. The assessed trabecular pattern is enclosed within a trapezoid marker symbolising ROI (Figure 2) that needs to be manually placed in order to perform the analysis. The marker is fixed in size and shape but could be moved and rotated. Only pixel data within the area of the marker are considered for the analysis. The area within the marker is resampled in size to make the resulting image match a fixed reference size. A novel technology has been introduced to the software that fully automates the whole process from placement of the trapezoid marker to the analysis of the trabecular pattern within. In other words, the fully automated tool removes the method's dependency on its user. The details of this technology are a corporate secret and cannot be included in this paper. The final resulting value represents the sum of the sizes and intensities of the spaces between the trabeculae. Values are between approximately 3000 (dense bone structure) and 9500 (sparse bone structure, *i.e*. large gaps between trabeculae). Values higher than 6500 denoted risk of osteoporosis according to the manufacturer's manual.

Analyses

Five raters performed analyses of the intraoral images using semi-automated software. Of the five raters one is a specialist in dental and maxillofacial radiology with 30 years of experience. The second rater was a general dental practitioner with 6 years of clinical experience and a postdoc in dental and maxillofacial radiology. The third rater was a general dental practitioner with 5 years of clinical experience and a PhD student. The last two raters were recently graduated general dental practitioners with less than 1 year of clinical experience. In addition to the analysis performed by the raters, the images were subjected to analysis with a fully automated tool provided by the most recent version of the software.

Prior to the analyses, an information session on the purpose of the study and a calibration exercise took place with all the raters. The assessment instructions were specified both verbally and in writing. All raters were familiar with the handling of the software.

The images were uploaded to the software in JPEG (Joint Photographic Experts Group), TIFF (Tagged Image File Format) or BMP format (Bitmap image file format; 8-bit greyscale). A trapezoid marker symbolising ROI was manually placed with the raters between the roots/apices of the premolar area in the lower jaw (Figure 2). In the case of analysis by the fully automated tool the trapezoid marker was placed by the software itself. The trabecular pattern was assessed by the algorithm provided in the software.

All raters performed the analyses independently of each other at their own location using computer screens with resolution 1920 × 1080. The observation rooms were dimly lit as recommended by the American Association of Physicists in Medicine Task Group.[28] The distance to the screen was approximately 50 cm. There was no restriction regarding observation time. All raters were blinded to clinical features such as the patients' age, previous medical history and individual DXA results. Software developed for the purpose performed a random selection of 121 images (20%) for the second round of observations performed by all raters, except for the fully automated tool, after 4 weeks.

Statistical methods

Rater agreement for categorical data was established through κ value using SPSS software v. 24.0 for Windows

(IBM Corp., Armonk, NY). Pairwise interrater agreement was calculated to define the reliability between the investigator's evaluation for each case. Intrarater agreement was calculated based on the reliability of the individual investigators between the first and the second analysis. Fleiss' κ was used to calculate the rater agreement between all the raters with and without the fully automated tool. Fleiss' κ was calculated using STATA software v. 16 for Windows (StataCorp LLC, College Station, TX). Sensitivity, specificity, predictive values as well as likelihood ratios and accuracy were calculated using the online statistical calculator MedCalc®(MedCalc Software, Ostend, Belgium).[29] To examine differences between groups for dichotomised variables we used $\chi 2$ test calculated using Chi-Square Calculator.[30]

Results

In this study, 567 radiographs that met the inclusion criteria were selected through consensus reached between the two authors. One radiograph was chosen for each participant.

The results acquired from the analyses were dichotomised into low and high risk of osteoporosis with a threshold value of 6500 units in agreement with the manufacturer's manual. The distribution of raters and the fully automated tool assessments into two groups according to the subjects' risk for osteoporosis is presented in Table 1. The number of subjects assessed as having high risk of osteoporosis ranged between 52 and 211. There was a difference between the raters in the number of radiographs excluded due to being deemed not possible to assess. Rater 1 did not exclude any radiographs, while Rater 4 excluded 149 out of 567 radiographs.

Tables 2 and 3 present κ values for pairwise inter- and intrarater agreement, respectively. Following Landis and Koch[31] interpretation scale, pairwise agreement between the five raters varies from fair to moderate. Agreement between the five raters and the softwares' fully automated tool is mostly fair. Intrarater agreement varies from moderate to substantial.[31] Overall interrater agreement κ values for the five raters was 0.37 (95% CI 0.32–0.41), and for the five raters with the fully automated tool included as a "sixth rater", the overall κ was 0.34 (0.30–0.38).

Table 1 Distribution of assessments dichotomised into low and high risk of osteoporosis with the threshold value of 6500 units for five raters and the fully automated tool (Auto.tool)

Rater	*Low risk of osteoporosis (<6500)*	*High risk of osteoporosis (≥6500)*	*Missing values*
1	423	144	0
2	446	52	69
3	351	211	5
4	255	163	149
5	332	201	34
Auto.tool	290	204	73

Figure 3, Table 4 show the diagnostic accuracy of osteoporosis risk assessment by five raters using the software as well as for the fully automated tool. The diagnostic accuracy was presented as means of specificity, sensitivity, predictive values, likelihood and odds ratio. The analysis was performed on dichotomised data with the threshold value of 6500 units in agreement with the manufacturers' manual. Accuracy was calculated in relation to reference standard for osteoporosis diagnosis which is T-score values for spine, total hip and femoral neck, respectively where values lower than −2.5 were denoted osteoporotic. With values higher than 6500 units as an indicator of high risk of osteoporosis in intraoral radiography, the mean sensitivity of all raters' including the fully automated tool was 40,4% (range 14,3–57,6%). The corresponding value for specificity was 69,5% (range 59,7–90,4%). There are large differences between the five raters. Rater 2 stood out in comparison to the others with low sensitivity values (18.4%, 22.6% and 14. 3% for T-score spine, T-score total hip and T-score femoral neck, respectively) and high specificity (90.4%, 90.4% and 90.1%). The fully automated tool sensitivity varies from 45.6% for T-score spine, 57.6% for T-score total hip to 49.3% for T-score femoral neck. Specificity of the fully automated tool for T-score spine was 68.8% which was higher than its specificity for T-score total hip and femoral neck (59.7 and 59.9%). The diagnostic odds ratios vary between 1 and 2.7.

Because the interval for collection of radiographs was relatively wide (±3 years from DXA examination) in order to control for the possibility of influence on the bone tissue and thus the results of the analysis, we performed a $\chi 2$ test for the analysis between groups based on their previous history of medication with antiosteoporotic drugs. We investigated if there was a difference between the group of study participants who either had been treated with osteoporosis medication before the study or at the start of the study with the group of study participants who never underwent osteoporosis treatment. The results were only statistically significant in two cases of specificity for two raters (Rater 2 and Rater 5) and the fully automated tool. There was no significant difference in sensitivity between the groups.

Discussion

The findings of this study show that even when a computer-based tool for assessment of trabecular bone pattern is used, there is variation among raters regarding where to apply the tool and subsequently the result of the assessments varies. We chose to evaluate one computer-based tool that is currently marketed to general dental practitioners as a way of identifying individuals with osteoporosis or at risk of developing osteoporosis.

Table 2 Pairwise interrater agreement calculated as unweighted κ for five raters and the fully automated tool (Auto.tool) with 95% CI

κ	*Rater 2*	*Rater 3*	*Rater 4*	*Rater 5*	*Auto.tool*
Rater 1	0,29 (0,19–0,38)	0,39 (0,31–0,46)	0,42 (0,33–0,51)	0,44 (0,36–0,51)	0,29 (0,21–0,37)
Rater 2		0,21 (0,14–0,28)	0,24 (0,16–0,31)	0,25 (0,18–0,32)	0,16 (0,09–0,23)
Rater 3			0,41 (0,32–0,50)	0,46 (0,38–0,53)	0,28 (0,20–0,37)
Rater 4				0,52 (0,44–0,61)	0,32 (0,23–0,42)
Rater 5					0,39 (0,31–0,48)

CI, confidence interval.

The software algorithm used in this study can be classified as artificial intelligence (AI). Although there is no straightforward definition, AI is considered to be the capability of a machine to imitate intelligent human behaviour.[32] Image processing and computer vision are both examples of AI applied in the field of medicine. Image processing is defined as a mathematical process enhancing an image to retrieve specific information like pattern measurements. Computer vision is defined as the processing of an image to enable identification of the image input and to provide an appropriate output.[33] Boneprox software algorithm falls perfectly under these definitions. However, its limitation lies in the fact that it has been developed and trained on the same small set of radiographs. While there are techniques that have been proved to successfully fine-tune a softwares' capability of recognising specific patterns using a limited amount of data,[34] the results of this study show that the software for trabecular pattern analysis in intraoral radiographs requires further work. Possibly Boneprox algorithm could improve its skills by training on a larger, more versatile volume of images and using a convolutional neural network (CNN), another form of AI. CNN has yielded impressive results in diagnosis and prediction within other fields of odontology,[35] such as diagnosing caries or periapical lesions.[36–38]

It has been suggested that the digitalisation of dental offices, including dental radiography, might lead to the development of software for trabecular pattern analysis on a large scale. Such a tool could be well suited for identifying individuals with or at risk of developing osteoporosis performed on a large number of radiographs already collected for diagnostic reasons.[39] Several digital image processing techniques have been designed to measure and describe the structure of the trabecular pattern. Algorithms based on a binary, skeletonised version of the image have been constructed for the sole purpose of analysing morphological variables of trabecular bone tissue, including trabecula and the intertrabecular spaces. These structures were chosen as they were considered to be related to each other (together they fill up the image) but with different properties that could influence the outcome of the analysis. Software were applied on among others iliac crest,[40,41] femoral

Table 3 Intrarater agreement calculated as unweighted κ for five raters with 95% CI

Rater	*Intrarater agreement*
1	0,57 (0,40–0,73)
2	0,73 (0,48–0,98)
3	0,49 (0,33–0,66)
4	0,70 (0,53–0,87)
5	0,65 (0,50–0,80)

CI, confidence interval.

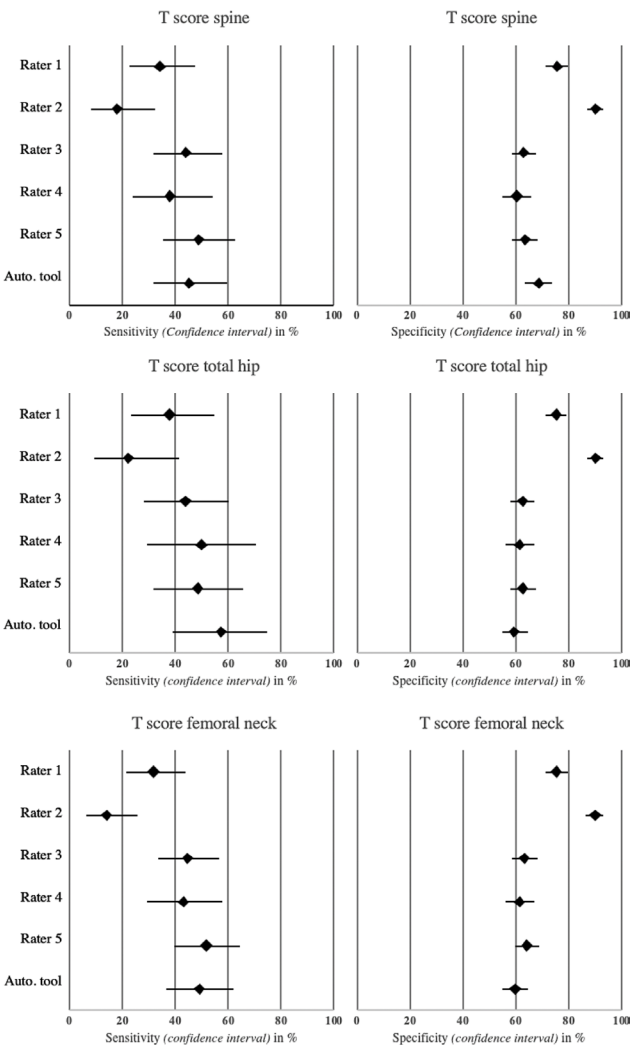

Figure 3 Sensitivity and specificity for five raters and the fully automated tool (Auto.tool) using the cut-off value of 6500 units indicating osteoporosis.

Table 4 Diagnostic accuracy for five raters and the fully automated tool (Auto.tool) for osteoporosis risk assessment in intraoral radiographs calculated as sensitivities, specificities, positive and negative predictive values with 95% CI and positive and negative likelihood ratios and odds ratio

T-score	*Rater*	*Sensitivity [%]*		*Specificity [%]*		*Predictive values [%]*				*Likelihood ratio*		*Odds ratio*
		Value	95% CI	Value	95% CI	Positive	95% CI	Negative	95% CI	Positive	Negative	
Spine	1	34,4	23,0–47,3	75,9	71,9–79,5	15,4	11,1–20,9	90,1	88,3–91,6	1,4	0,9	1,6
	2	18,4	8,8–32,0	90,4	87,3–93,0	17,3	9,8–28,7	91,0	89,8–92,1	1,9	0,9	2,1
	3	44,4	31,9–57,5	63,4	59,0–67,6	13,3	10,2–17,2	90,0	87,7–91,9	1,2	0,9	1,4
	4	38,3	24,5–53,6	60,8	55,6–65,8	11,0	7,8–15,4	88,6	85,9–90,8	1,0	1,0	1,0
	5	49,2	35,9–62,5	63,8	59,3–68,1	14,5	11,3–18,4	91,0	88,6–92,9	1,4	0,8	1,7
	Auto.tool	45,6	32,4–59,3	68,8	63,8–73,5	18,1	13,9–23,5	89,3	86,7–91,4	1,5	0,8	1,9
Hip	1	38,1	23,6–54,4	75,6	71,7–79,2	11,1	7,6–15,9	93,8	92,3–95,1	1,6	0,8	1,9
	2	22,6	9,6–41,1	90,4	87,3–92,9	13,5	7,1–24,0	94,6	93,6–95,5	2,3	0,9	2,7
	3	43,9	28,5–60,3	62,9	58,6–67,1	8,5	6,1–11,8	93,4	91,5–95,0	1,2	0,9	1,3
	4	50,0	29,9–70,1	61,7	56,7–66,6	8,0	5,5–11,5	94,9	92,6–96,5	1,3	0,8	1,6
	5	48,7	31,9–65,6	63,0	58,6–67,3	8,96	6,5–12,3	94,3	92,3–95,8	1,3	0,8	1,6
	Auto.tool	57,6	39,2–74,5	59,7	55,1–64,3	9,3	7,0–12,3	95,2	92,9–96,7	1,4	0,7	2,0
Femoral neck	1	32,1	22,2–43,4	75,7	71,6–79,4	18,1	13,4–23,9	87,0	85,1–88,7	1,3	0,9	1,5
	2	14,3	6,8–25,4	90,1	86,9–92,8	17,3	9,7–29,0	88,0	86,7–89,0	1,5	1,0	1,5
	3	45,0	33,9–56,5	63,6	59,1–67,9	17,1	13,6–21,2	87,4	84,9–89,6	1,2	0,9	1,4
	4	43,4	29,8–57,7	61,6	56,4–66,7	14,1	10,5–18,7	88,2	85,4–90,6	1,1	0,9	1,2
	5	52,1	40,0–63,9	64,5	59,9–68,9	18,9	15,3–23,1	89,4	86,8–91,6	1,5	0,7	2,0
	Auto.tool	49,3	36,8–61,8	59,9	55,1–64,6	16,3	12,9–20,3	88,2	85,4–90,6	1,2	0,9	1,5

CI, confidence interval.

neck,[42] vertebrae[43] and distal radius.[44,45] An equivalent method was applied on intraoral periapical radiographs in search for signs of osteoporosis showing architectural changes in the trabecular bone in osteoporotic individuals in comparison to healthy counterparts.[12,17,18,46] Within the jaws, the premolar region of the lower jaw has been of particular interest due to little variation in anatomy and the lack of major muscular fibre insertion in this region.[47] Another advantage of this region is that when no teeth are present, the mental foramen serves as a landmark for the region to assess. However, the method applied on intraoral radiographs was never taken to the next step of being tested in a large-scale population study. Another approach of automated analysis of trabecular pattern called Jax-X method, was also applied on intraoral radiographs with focus on identification of the analysis of intertrabecular spaces instead.[48,49] The software investigated in this study is the next generation of the Jaw-X method. It is difficult to establish in detail what the similarities or differences between the software are since the algorithm is either developed and applied locally or protected by patent or trade secret.

Another issue with previous studies is that the study samples were usually small, with the smallest sample consisting of 23 individuals.[12] Only one study presented a sample size big enough (=671) to draw a reliable conclusion.[18] We intended to analyse images from a big sample, which is why we decided to select images from patients taking part in the SUBERB study[27] with 3028 participants. Due to difficulties in obtaining images from the clinics and image quality in many cases being too poor to allow analyses, the final number of radiographs available for analysis was 567 from the same number of patients, a sample size still superior to that used in several other studies.

On one hand, a prerequisite for any method, visual or computer-based, is that image quality regarding projection and exposure settings is close to optimal. In well-planned prospective research projects, image quality is more likely to be optimal. On the other hand, a previous study has shown that many of the morphologic variables of trabecular bone are robust enough to withstand variation in exposure angle and image brightness (optical density).[39] With this being our hypothesis and the fact that the software under investigation in this study is already being marketed to general dental practitioners in Sweden, we concluded that the method should be tested on radiographs obtained during routine clinical work in general dental practice. Therefore, we analysed radiographs collected retrospectively and taken in ordinary clinical settings.

It is tempting to believe that if using a computer-based tool, the influence of its user will be eliminated. However, unless fully automated, the impact of human interaction cannot be neglected. In this study, we used five raters to explore any differences in results between and within raters when using a computer-based digital

image processing software. In addition, we applied a fully automated version of that software that was to be used without human interaction. When, for any reason, analyses or measurements are performed in radiographs, both agreement and reliability are important concepts as they provide information about the quality of measurements.[23]

To the best of our knowledge, no other study has compared rater agreement using semi-automated software for trabecular pattern analysis in dental radiographs. The degree of agreement between the raters in this study being only fair to moderate is comparable with the agreement between raters in previous studies using visual assessment method to assess trabecular pattern.[13] Moreover, the degree of experience and expertise of raters in this study varied considerably, whilst in the study using a visual assessment approach, four out of five raters were highly trained in the field of oral radiology and had more than 20 years of experience. Additionally, the image acquisition in that study was more standardised and obtained in the controlled environment of specialist clinics. With that in mind, our study showed that the evaluated software achieved similar levels of inter- and intrarater agreement used on radiographs attained retrospectively from unknown conditions and analysed by raters with different levels of clinical experience. However, the fully automated tool built into the software to eliminate the user's influence showed only slight to fair agreement with the five raters, which indicates it needs further improvement before it can be recommended for commercial use.

Software accuracy has also been investigated through sensitivity and specificity analysis in relation to participants T-score values in spine, hip and femoral neck. In general, sensitivity was lower than the specificity for most raters and the fully automated tool. This agrees with previous studies investigating the same software sensitivity and specificity.[48] The fully automated tools' sensitivity varied between 45.6 and 57.9% and specificity between 59.8 and 68.8%, which means that its ability to identify individuals with risk of osteoporosis and individuals without risk of osteoporosis is almost equal. In other words, the fully automated tool would generate many false-positive and false-negative results at the general dental clinic. False-positive results are unwanted because it not acceptable to worry dental patients unnecessarily and refer them for DXA measurements. In contrast, false-negative results keep the patients unaware of a condition they might have that may lead to serious health issues such as fragility fracture.

A number of strengths and limitations of this study must be considered. One limitation of this study was the difficulty of collecting radiographs from general dental practitioners. We tried to retrieve images from general dental practices by contacting the clinics directly, however, with limited outcome. When using retrospectively collected images from different clinics, the information regarding image acquisition is limited. Different technologies for image acquisition (analogue radiographs scanned, digital radiographs) and various file formats (DICOM converted to JPEG, JPEG, TIFF and BMP) could also be a potential source of bias. Inclusion criteria for the radiographs followed previously established image quality criteria as well as the softwares' general requirements regarding ROI. Although clear quality criteria had been set before the assessments, there was a difference between the raters in how they judged the quality of a single radiograph to be sufficient for assessment. On one hand, this could be perceived as another source of bias, but it should also be seen as a strength of this study as it reflects the reality of general dental practice. Another limitation is that it was not possible to investigate the reproducibility of the fully automated tool. This tool could only be handled by the manufacturer and despite several inquiries we could not get the assistance.

The strength of this study is that it was part of a large population study which made it possible to acquire a large amount of material for analysis compared to many previous studies. Also, unique for this study is that the cohort (SUPERB) had a rather narrow age span (75–80 years) and that all participants underwent DXA examination, which enabled us to evaluate the accuracy of the software.

Conclusions

The low diagnostic odds ratio and agreement between raters in osteoporosis risk assessment using the investigated software leads to the conclusion that more work needs to be done to optimise the automation process of analysis of the trabecular pattern in intraoral radiographs.

Acknowledgements

Ingvar Kamprad Foundation, C O Henriksson Foundation and Swedish Dental Association financially supported the research for this paper. There is no conflict of interest. We gratefully acknowledge technical support from Michael Gullberg.

Funding

Ingvar Kamprad Foundation, C O Henriksson Foundation, Swedish Dental Association

REFERENCES

1. Peck WA, Burckhardt P, Christiansen C, Fleisch HA, Genant HK, Gennari C. Consensus development conference: diagnosis, prophylaxis, and treatment of osteoporosis. *Am J Med* 1993; **94**: 646–50. doi: https://doi.org/10.1016/0002-9343(93)90218-e
2. Sambrook P, Cooper C. Osteoporosis. *Lancet* 2006; **367**: 2010–8. doi: https://doi.org/10.1016/S0140-6736(06)68891-0
3. Hernlund E, Svedbom A, Ivergård M, Compston J, Cooper C, Stenmark J, et al. Osteoporosis in the European Union: medical management, epidemiology and economic burden. A report prepared in collaboration with the International osteoporosis Foundation (IOF) and the European Federation of pharmaceutical industry associations (EFPIA. *Arch Osteoporos* 2013; **8**: 136. doi: https://doi.org/10.1007/s11657-013-0136-1
4. Kanis JA, Johnell O, Oden A, Sembo I, Redlund-Johnell I, Dawson A, et al. Long-Term risk of osteoporotic fracture in Malmö. *Osteoporos Int* 2000; **11**: 669–74. doi: https://doi.org/10.1007/s001980070064
5. WHO Assessment of fracture risk and its application to screening for postmenopausal osteoporosis. Report of a who Study Group. *World Health Organ Tech Rep Ser* 1994; **843**: 1–129.
6. Kanis JA, Cooper C, Rizzoli R, Reginster JY. Scientific Advisory Board of the European Society for C, economic aspects of O, et al. European guidance for the diagnosis and management of osteoporosis in postmenopausal women. *Osteoporos Int* 2019; **30**: 3–44.
7. SBU.*Osteoporosis - prevention, diagnosis and treatment*. Stockholm: Swedish Council on Health Technology Assessment in Health Care; 2003.
8. Horner K, Devlin H. Clinical bone densitometric study of mandibular atrophy using dental panoramic tomography. *J Dent* 1992; **20**: 33–7. doi: https://doi.org/10.1016/0300-5712(92)90007-y
9. Horner K, Devlin H, Alsop CW, Hodgkinson IM, Adams JE. Mandibular bone mineral density as a predictor of skeletal osteoporosis. *Br J Radiol* 1996; **69**: 1019–25. doi: https://doi.org/10.1259/0007-1285-69-827-1019
10. Taguchi A, Tanimoto K, Suei Y, Ohama K, Wada T. Relationship between the mandibular and lumbar vertebral bone mineral density at different postmenopausal stages. *Dentomaxillofac Radiol* 1996; **25**: 130–5. doi: https://doi.org/10.1259/dmfr.25.3.9084261
11. Lindh C, Obrant K, Petersson A. Maxillary bone mineral density and its relationship to the bone mineral density of the lumbar spine and hip. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2004; **98**: 102–9. doi: https://doi.org/10.1016/s1079-2104(03)00460-8
12. White SC, Rudolph DJ. Alterations of the trabecular pattern of the jaws in patients with osteoporosis. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 1999; **88**: 628–35. doi: https://doi.org/10.1016/s1079-2104(99)70097-1
13. Lindh C, Horner K, Jonasson G, Olsson P, Rohlin M, Jacobs R, et al. The use of visual assessment of dental radiographs for identifying women at risk of having osteoporosis: the OSTEODENT project. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2008; **106**: 285–93. doi: https://doi.org/10.1016/j.tripleo.2007.09.008
14. Jonasson G, Alstad T, Vahedi F, Bosaeus I, Lissner L, Hakeberg M. Trabecular pattern in the mandible as bone fracture predictor. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2009; **108**: e42–51. doi: https://doi.org/10.1016/j.tripleo.2009.05.018
15. Norderyd O, Kochi G, Papias A, Kohler AA, Helkimo AN, Brahm CO. Oral health of individuals aged 3-80 years in Jonkoping, Sweden, during 40 years (1973-2013). *I. Review of findings on oral care habits and knowledge of oral health. Swed Dent J* 2015; **39**: 57–68.
16. Waite S, Scott J, Gale B, Fuchs T, Kolla S, Reede D. Interpretive error in radiology. *Am J Roentgenol Radium Ther* 2017; **208**: 739–49.
17. White SC, Cohen JM, Mourshed FA. Digital analysis of trabecular pattern in jaws of patients with sickle cell anemia. *Dentomaxillofac Radiol* 2000; **29**: 119–24. doi: https://doi.org/10.1038/sj/dmfr/4600516
18. Verheij JGC, Geraets WGM, van der Stelt PF, Horner K, Lindh C, Nicopoulou-Karayianni K, et al. Prediction of osteoporosis with dental radiographs and age. *Dentomaxillofac Radiol* 2009; **38**: 431–7. doi: https://doi.org/10.1259/dmfr/55502190
19. Licks R, Licks V, Ourique F, Radke Bittencourt H, Fontanella V. Development of a prediction tool for low bone mass based on clinical data and periapical radiography. *Dentomaxillofac Radiol* 2010; **39**: 224–30. doi: https://doi.org/10.1259/dmfr/23760876
20. Geraets WGM, Lindh C, Verheij H. Sparseness of the trabecular pattern on dental radiographs: visual assessment compared with semi-automated measurements. *Br J Radiol* 2012; **85**: e455–60. doi: https://doi.org/10.1259/bjr/32962542
21. Geraets WGM, Verheij JGC, van der Stelt PF, Horner K, Lindh C, Nicopoulou-Karayianni K, et al. Selecting regions of interest on intraoral radiographs for the prediction of bone mineral density. *Dentomaxillofac Radiol* 2008; **37**: 375–9. doi: https://doi.org/10.1259/dmfr/29966973
22. Farzin B, Gentric J-C, Pham M, Tremblay-Paquet S, Brosseau L, Roy C, et al. Agreement studies in radiology research. *Diagn Interv Imaging* 2017; **98**: 227–33. doi: https://doi.org/10.1016/j.diii.2016.05.014
23. Kottner J, Gajewski BJ, Streiner DL. Guidelines for reporting reliability and agreement studies (GRRAS). *Int J Nurs Stud* 2011; **48**: 659–60. doi: https://doi.org/10.1016/j.ijnurstu.2011.01.017
24. Larsson B, Mellström D, Johansson L, Nilsson AG, Lorentzon M, Sundh D. Normal bone microstructure and density but worse physical function in older women treated with selective serotonin reuptake inhibitors, a cross-sectional population-based study. *Calcif Tissue Int* 2018; **103**: 278–88. doi: https://doi.org/10.1007/s00223-018-0427-z
25. Nilsson AG, Sundh D, Johansson L, Nilsson M, Mellström D, Rudäng R, et al. Type 2 diabetes mellitus is associated with better bone microarchitecture but lower bone material strength and poorer physical function in elderly women: a population-based study. *J Bone Miner Res* 2017; **32**: 1062–71. doi: https://doi.org/10.1002/jbmr.3057
26. Looker AC, Wahner HW, Dunn WL, Calvo MS, Harris TB, Heyse SP, et al. Updated data on proximal femur bone mineral levels of US adults. *Osteoporos Int* 1998; **8**: 468–89. doi: https://doi.org/10.1007/s001980050093
27. Lorentzon M, Nilsson AG, Johansson H, Kanis JA, Mellström D, Sundh D. Extensive undertreatment of osteoporosis in older Swedish women. *Osteoporos Int* 2019; **30**: 1297–305. doi: https://doi.org/10.1007/s00198-019-04872-4
28. Samei E, Badano A, Chakraborty D, Compton K, Cornelius C, Corrigan K, et al. Assessment of display performance for medical imaging systems: Executive summary of AAPM TG18 report. *Med Phys* 2005; **32**: 1205–25. doi: https://doi.org/10.1118/1.1861159
29. MedCalc.Diagnostic test evaluation calculator. 2021. Available from: https://www.medcalc.org/calc/diagnostic_test.php.
30. Stangroom J. Social Science Statistics. 2018. Available from: https://www.socscistatistics.com/tests/chisquare/default2.aspx.
31. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–74.
32. Webster M. Definition: 'Artificial Intelligence. Available from: https://www.merriam-webster.com/dictionary/artificial%20intelligence.
33. Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. *Minim Invasive Ther Allied Technol* 2019; **28**: 73–81. doi: https://doi.org/10.1080/13645706.2019.1575882
34. Lee K-S, Jung S-K, Ryu J-J, Shin S-W, Choi J. Evaluation of transfer learning with deep Convolutional neural networks for screening osteoporosis in dental panoramic radiographs. *J Clin Med* 2020; **9**: 392. doi: https://doi.org/10.3390/jcm9020392

35. Schwendicke F, Golla T, Dreher M, Krois J. Convolutional neural networks for dental image diagnostics: a scoping review. *J Dent* 2019; **91**: 103226. doi: https://doi.org/10.1016/j.jdent.2019.103226
36. Lee J-H, Kim D-H, Jeong S-N, Choi S-H. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent* 2018; **77**: 106–11. doi: https://doi.org/10.1016/j.jdent.2018.07.015
37. Pauwels R, Brasil DM, Yamasaki MC, Jacobs R, Bosmans H, Freitas DQ, et al. Artificial intelligence for detection of periapical lesions on intraoral radiographs: comparison between convolutional neural networks and human observers. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2021; **131**: 610–6. doi: https://doi.org/10.1016/j.oooo.2021.01.018
38. Karatas O, Cakir NN, Ozsariyildiz SS, Kis HC, Demirbuga S, Gurgan CA. A deep learning approach to dental restoration classification from bitewing and periapical radiographs. *Quintessence Int* 2021; **52**: 568–74. doi: https://doi.org/10.3290/j.qi.b1244461
39. White S. Influence of X-ray beam angulation and exposure on morphologic features of trabecular bone. *Int J Oral Biol* 1999; **24**: 17–23.
40. Compston JE, Mellish RW, Garrahan NJ. Age-Related changes in iliac crest trabecular microanatomic bone structure in man. *Bone* 1987; **8**: 289–92. doi: https://doi.org/10.1016/8756-3282(87)90004-4
41. Croucher PI, Garrahan NJ, Compston JE. Structural mechanisms of trabecular bone loss in primary osteoporosis: specific disease mechanism or early ageing? *Bone Miner* 1994; **25**: 111–21. doi: https://doi.org/10.1016/s0169-6009(08)80253-x
42. Geraets WG, Van der Stelt PF, Lips P, Van Ginkel FC. The radiographic trabecular pattern of hips in patients with hip fractures and in elderly control subjects. *Bone* 1998; **22**: 165–73. doi: https://doi.org/10.1016/s8756-3282(97)00233-0
43. Caligiuri P, Giger ML, Favus MJ, Jia H, Doi K, Dixon LB. Computerized radiographic analysis of osteoporosis: preliminary evaluation. *Radiology* 1993; **186**: 471–4. doi: https://doi.org/10.1148/radiology.186.2.8421753
44. Geraets WG, Van der Stelt PF, Netelenbos CJ, Elders PJ. A new method for automatic recognition of the radiographic trabecular pattern. *J Bone Miner Res* 1990; **5**: 227–33. doi: https://doi.org/10.1002/jbmr.5650050305
45. Geraets WG, Van der Stelt PF, Elders PJ. The radiographic trabecular bone pattern during menopause. *Bone* 1993; **14**: 859–64. doi: https://doi.org/10.1016/8756-3282(93)90315-2
46. Lee BD, White SC. Age and trabecular features of alveolar bone associated with osteoporosis. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2005; **100**: 92–8. doi: https://doi.org/10.1016/j.tripleo.2004.11.020
47. von Wowern N, Stoltze K. Histoquantitation on small jaw specimens. *Scand J Dent Res* 1978; **86**: 193–9. doi: https://doi.org/10.1111/j.1600-0722.1978.tb01931.x
48. Hassani-Nejad A, Ahlqwist M, Hakeberg M, Jonasson G. Mandibular trabecular bone as fracture indicator in 80-year-old men and women. *Eur J Oral Sci* 2013; **121**: 525–31. doi: https://doi.org/10.1111/eos.12087
49. Jonasson G, Billhult A. Mandibular bone structure, bone mineral density, and clinical variables as fracture predictors: a 15-year follow-up of female patients in a dental clinic. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2013; **116**: 362–8. doi: https://doi.org/10.1016/j.oooo.2013.06.009