UiT The Arctic University of Norway

Faculty of Health Sciences

Department of Clinical Dentistry

Panoramic radiograph analyses for early detection of osteoporosis in the population of Northern Norway

Anna Teterina

A dissertation for the degree of Philosophiae Doctor, October 2023

Panoramic radiograph analyses for early detection of osteoporosis in the population of Northern Norway.

Anna Teterina

A dissertation for the degree of Philosophiae Doctor, October 2023

Department of Clinical Dentistry, Faculty of Health Sciences,

University of Tromsø–The Arctic University of Norway, Tromsø, Norway

Table of Contents

List of Figures

Foreword

This thesis was carried out at the Department of Clinical Dentistry (IKO), Faculty of Health Sciences, UiT, The Arctic University of Norway. I am grateful to the Arctic University of Norway for training me as a researcher, initially through the Master of Public Health program and subsequently during the Ph.D. project in oral radiology. I also want to express my gratitude to all who have supported me throughout this project.

First, I am deeply grateful to my principal supervisor, associate professor at UiT Napat L. Bolstad. She introduced me to the oral radiology research community and offered a lot of educational activities. She was always there for me, being ready to talk and help. I am incredibly grateful to her for arranging our collaboration with the Machine Learning Group at UiT and introducing artificial intelligence to me. It sparked my interest in artificial intelligence and its use in medical imaging.

I express my sincere appreciation to my co-supervisor, Sanyalak Niratisairak, working at the Department of Orthopedics, Institute for Clinical Medicine, University of Oslo, who has been a great help in explaining to me the basics of digital images and technicalities of DEXA scans, thoroughly reading all my drafts, giving valuable feedback, making great graphs and jokes.

I am also grateful to my co-supervisor, Professor at UiT Bente Morseth, for her significant contribution to this project. She generously shared her expertise in osteoporosis studies and epidemiology, providing valuable insights and feedback on my work.

Moreover, I want to thank my collaborators and co-authors in this project, Isak Paasche Edvardsen, Thomas Johansen, Jonas Nordhaug Myhre, and Professor Fred Godtliebsen. Working with them and getting familiar with machine learning research was fascinating.

I want to express my gratitude to my colleagues at IKO, where I have spent four years and have been involved in many different activities. First, I am deeply thankful to Professor Hans-Peter Műller, who is now on retirement, for his professional advice and exciting conversations that made me feel smart. I would also like to thank Glennda Villaflor and Rania Al-Mahdi for having fun chatting and organizing social events at IKO. Supervising dental hygienists in their bachelor thesis was another valuable experience for me, and I am thankful to Linda Stein, Kine Mathisen, Silje Navjord Moltubakk, and Hilde Nyborg for giving me this opportunity. I also thank Laila, Kristin, Inger-Anna, Rita, Anders, Gro Eirin, Bo, and others for friendly chats in the lunch room. I appreciate it a lot. A special mention goes to my former colleagues Kaisa and Jukka Leinonen, who have evolved into family friends, and I miss them here in Tromsø. I am also very grateful to Lina Stangvaltaite-Mouhat for being my job reference and a friendly colleague.

I thank my dear friends Liubov, Alexsander, Irina, Stanislav, Yana, and Ilia for supporting me in friendly chats and meetings. I am deeply grateful to my parents, Tatiana, Sergei, and my brother Viacheslav, for their unconditional love and cheering on me. I also thank my parents-in-law, Inna and Andrey, and the Valkov family for caring. My darling husband, Sergei, deserves a special mention for his endless love, for believing in me, and for supporting me in whatever I do. Last but certainly not least, I want to express my gratitude to my cherished son, Roman. His warm hugs, thoughtful questions about everything, and the joy we shared playing PC games and relaxing together have enriched my life.

Thank you all for being part of this journey!

Abbreviations and terms

DEXA – dual energy x-ray absorptiometry

BMD – bone mineral density

FRAX – fracture risk assessment tool

OSIRIS – osteoporosis index risk

BMI – body mass index

BMC – bone mineral content

MCW – mandibular cortical width

MCI – mandibular cortical index

DPR – dental panoramic radiograph

ROI – region of interest

MF – mental foramen

PPD – periodontal probing depth

BOP – bleeding on probing

RBL -radiological bone level

CI – confidence intervals

PPV – positive predictive value

NPV – negative predictive value

LR+ – positive likelihood ratio

LR- – negative likelihood ratio

ROC – receiver operating characteristic curve

AUC – area under ROC curve

OR – odds ratio

mAP – mean average precision (corresponds to AUC)

ASM – active shape models

List of papers

The thesis is based on the following papers:

1. Diagnostic efficacy of radiomorphometric indices for predicting osteoporosis in a Norwegian population in the Tromsø Study: Tromsø7.

Oral Surgery, Oral Medicine, Oral Pathology, and Oral Radiology. 2023; 135(3); 444-55.

Doi: 10.1016/j.oooo.2022.10.039

Teterina A, Niratisairak S, Morseth B, Bolstad N.

2. General and local predictors of mandibular cortical bone morphology in adult females and males: The seventh survey of the Tromsø Study.

Clinical Oral Investigations. 2023

Doi: 10.1007/s00784-023-05263-0

Teterina A, Niratisairak S, Morseth B, Bolstad N.

3. **Automatic detection of the mental foramen for estimating mandibular cortical width in dental panoramic radiographs: the seventh survey of the Tromsø Study (Tromsø7) in 2015–2016.**

Journal of International Medical Research. 2022; 50(11).

Doi: 10.1177/03000605221135147.

Paasche Edvardsen I, Teterina A, Johansen T, Nordhaug Myhre J, Godtliebsen F, Limchaichana Bolstad N.

Summary

Osteoporosis is a chronic disease of bone tissue that increases the risk of fractures from minor accidents that would not cause fractures in healthy individuals. Roughly 20 % of females and 6% of males have osteoporosis after the age of 50, but the disease may also be present earlier in life. Osteoporosis exhibits no symptoms until the first fracture, and the early diagnosis is challenging. Dental radiography is a frequent examination, and the possibility of using it for osteoporosis screening has been studied for a few decades.

This thesis aims to investigate the diagnostic qualities of mandibular cortical shape and width assessments for osteoporosis screening in Norwegian males and females, to examine the factors affecting mandibular cortical morphology, and finally to explore the feasibility of automatic measurement of mandibular cortical width.

The thesis is based on the data from the seventh survey of the Tromsø study (Tromsø7) carried out in 2015-2016. Participants aged 40 and older were examined with dental panoramic radiographs and dual-energy x-ray absorptiometry at the femoral neck. Other demographic, health, and lifestyle data were collected in questionnaires.

Thin ($\leq$ 3 mm) and severely eroded cortex could differentiate osteoporotic from non-osteoporotic females $\geq$ 40, demonstrating sensitivities of 68.1% and 34.0% and specificities of 69.0% and 89.7%, respectively. Combining mandibular cortical width and shape with Fracture Risk Assessment (FRAX) score markedly improved their diagnostic efficacy with an increased ability to identify healthy females. T-score was the strongest predictor of mandibular cortical morphology in females, followed by age and number of remaining teeth. Only the T-score was weakly associated with cortical shape in males, while the efficacy estimates for radiomorphometric indices were inconclusive. The reproducibility of the manually measured width and shape was suboptimal. However, it was feasible to develop a

fully automated algorithm for measuring MCW according to the original method with the EfficientDet neural network, having the highest accuracy of 79% in identifying mental foramen.

To conclude, thin and eroded cortex on dental panoramic radiographs might be as useful as existing risk-factor-based tools for osteoporosis screening in females, and their combination with the FRAX score has superior diagnostic ability than radiomorphometric indices used alone. The performance of fully automated index measurements should be explored in further, more extensive studies.

1 Introduction

Osteoporosis is a systemic, chronic, non-communicable disease characterized by microstructural deterioration, mass reduction, and excessive fragility of bones. Osteoporosis increases the risk of fracture. Osteoporotic fractures occur mainly in the hip, wrist, or spine. They may result from a fall from standing height or other minor accidents that would not cause fractures in healthy people. Osteoporosis is also called a "silent disease" since it has no symptoms until the first osteoporotic fracture occurs. Although an early diagnosis of osteoporosis is challenging for many reasons, it is crucial since effective therapies preventing osteoporotic fractures exist [1].

Dental radiography is a frequent examination. A report on using radiographs among dental practitioners in Norway 2006 showed that they carried out around 7.29 million radiographic examinations, accounting for 64% of all medical radiographic examinations, while 4.7 million citizens lived in Norway in 2006 [2, 3]. Dental panoramic radiographs comprise around 2% of the total dental radiographs, i.e., approximately 144,000 a year [2]. The idea that dentists could identify people susceptible to osteoporosis using dental panoramic radiographs is rational because people regularly attend dental offices, and the radiographs taken for dental reasons show bone tissue besides tooth structures. Seventy-four percent of the adult population visited their dentists once a year in 2012; this attendance rate has been stable since 2002 [4].

Moreover, dental office attendance is more prevalent in aged people living in urban areas, specifically females [5, 6]. In Norway, 77% of females attended dental offices once a year in 2012 compared to 70% of males; the highest percentage of dental attendance was in the age group of 45-66 in both sexes. A lower attendance percentage was observed in the age group

of 67 and older, but it was still higher than in the younger cohorts [4]. The abovementioned population attending dental clinics markedly coincides with the leading risk group for osteoporosis and osteoporotic fractures [7, 8].

1.1 Epidemiology and burden of osteoporosis

In the European Union, the estimated prevalence of osteoporosis in females and males aged 50 and older is 21% and 6%, respectively, yet some variations by country are observed [9]. The corresponding numbers in the United States are 19.6% and 4.4% for females and males [10]. A population-based study from Northern Norway revealed 11.5% of females and 8.3% of males aged 50 and older with osteoporosis in femoral necks [11]. The Norwegian Institute of Public Health estimated that around 300,000 out of 1.6 million Norwegians over 50 had osteoporosis in 2009 [12, 13].

The incidence of hip fracture varies remarkably across different countries, with the Scandinavian countries being among the world's top [14]. Osteoporosis causes around 9 million hip fractures, 1.7 million forearm fractures, and 1.4 million vertebral fractures annually worldwide [14]. In the E.U., 827,000 hip fractures, 663,000 vertebral, and 637,000 osteoporotic forearm fractures were sustained in 2019 [15]. According to the Norwegian Institute of Public Health, approximately 9,000 elderly persons experience a hip fracture, while 15,000 persons have forearm and vertebral fractures annually in Norway [12]. Seventy percent of all osteoporotic fractures occur in females. One-third of women and one-fifth of men over 50 will experience an osteoporotic fracture in their remaining lifetime [14].

Osteoporotic fractures increase mortality: around twenty percent die during the first year after hip fracture [16]. They also cause long-term disability, loss of independence, and chronic pain [14]. The impact of osteoporotic fractures in 27 countries of the European Union was estimated to be 112,850 disability-adjusted life years (DALYs), i.e., the sum of life years lost due to premature deaths and years lived with disability after sustaining an osteoporotic fracture, accounted for a weight value assigned to osteoporotic fracture [17]. In 2019, 56.9 billion euros were spent on medical care for patients with osteoporotic fractures in the European Union [17]. Since 2010, total medical care costs for osteoporotic fractures have increased by 64% [17]. Osteoporotic fractures are the fourth leading cause of chronic morbidity and disability in Europe after ischemic heart disease, dementia, and lung cancer [14, 17]. Norway spends 7-8 billion Norwegian kroner annually to treat hip fractures [12].

As the population ages, the osteoporotic fracture burden will increase [18]. By 2040, the number of people worldwide at high risk of hip fracture will double, while in Europe, this number is proposed to increase by 30% [18]. By 2050, one in five people in the Norwegian population will be older than 70 [19]. The report by Norwegian Epidemiologic Osteoporosis Studies (NOREPOS) from 2020 proposes that despite the 0.7% annual decline in osteoporotic fracture incidence, their absolute number will increase due to the growth and aging of the Norwegian population. The report also proposes doubling the osteoporotic fracture burden measured in DALYs and a 65% increase in medical care costs from 2020 until 2040 [20].

1.2 Clinical diagnosis of osteoporosis

Clinicians use dual-energy X-ray absorptiometry (DEXA) to diagnose osteoporosis. DEXA is an enhanced x-ray technology measuring bone mineral density (BMD) and a T-score. The T-score is a difference between the BMD of a patient and the mean BMD of young sex-matched

adults (reference population) divided by the standard deviation of the reference population [21]. The recommended reference population for calculating T-score is young females between 20 and 29 years old from the third National Health and Nutrition Examination Survey (NHANES III) [22]. Other sex-specific references are also standard options in different DEXA units [11, 23]. The operational definition of osteoporosis is a T-score assessed at the femoral neck equal to or less than the -2.5 standard deviation, while a T-score range between -1 to -2.5 standard deviations is considered osteopenia or low bone mineral density [22]. Osteopenia is not considered a disease and is primarily relevant for research. T-score is a strong predictor of osteoporotic fractures. The T-score at the hip has a higher predictive value for hip fracture than the other skeletal sites [22].

The main strengths of the technique are a relatively low effective dose of 1-6 μSv, high reproducibility, and ease of carrying out. However, the major limitation of DEXA is that it is a two-dimensional scan that measures an areal BMD. Even though sometimes bones have similar volumetric BMD, the areal BMD can be underestimated in small bones and overestimated in bigger ones because the third dimension, i.e., the depth, has not been considered. Moreover, DEXA is unreliable in persons with excessive weight changes, medical implants, spondylosis, or osteomalacia [24].

1.3 Other techniques assessing bone mass quantity

Radiogrammetry is the quantitative measurement of bone on radiographs, such as, for example, cortical thickness measured in the middle portion of the metacarpal bones called the "metacarpal index." This technique was revitalized in the 2000s when computer vision came. Quantitative computed tomography is another technique that gives a crossectional image of bones and allows the measurement of the volumetric bone mineral density and is not size-

dependent; it also allows measurements of cortical BMD separately from trabecular BMD. However, quantitative CT has a higher effective dose and is less available than DEXA. It is mainly used for research. Quantitative ultrasonography is also used for bone mass quantification. Unlike DEXA, the technique is based on short wavelengths and does not use ionizing radiation. However, the standard threshold for osteoporosis diagnosis does not fit, calibration between different devices is unavailable, and reliability is poor. Other high-resolution techniques also exist. However, they are less available, less studied, and sometimes use high-dose radiation (up to 3 mSv for high-resolution computed tomography) [24].

1.4 Screening for osteoporosis

An effective therapy preventing osteoporosis fractures exists [25, 26]. However, a populational-based screening for osteoporosis is not widely practiced. One reason is that the T-score measured by DEXA showed a suboptimal ability to predict osteoporotic fractures [7, 27]. Another reason is the low availability of DEXA in many world regions [7]. It has been estimated that 10.6 DEXA units per million of the general population are required [28]. The audit in different E.U. countries revealed that only 60% of 27 member states had the ~~r~~ecommended number of DEXA units [28]. In Northern Norway, only a few DEXA scanners exist for the entire population of Nordland, Troms, and Finnmark counties, as reported in 2020 [29].

Cost-effectiveness, i.e., gains in health relative to costs of osteoporosis screening using DEXA, would vary from country to country due to demographical, epidemiological, clinical practice, health care regulation, relative price difference, and other relevant factors [30]. Studies on osteoporosis screening cost-effectiveness are lacking. For example, no evidence of cost-effectiveness from the existing literature was concluded in Norwegian and Swedish

recommendations for medical doctors [31, 32]. However, both health technology assessments were carried out in the early 2000s and then not updated. On the contrary, screening and the following treatment were cost-effective compared to no screening in the United Kingdom [33]. In the U.S., all females older than 65 and males older than 70 with certain risk factors are recommended for osteoporosis screening with DEXA based on the cost-effectiveness of preventive treatment of this cohort [34, 35].

Since DEXA is not widely available for osteoporosis screening, case-finding strategies should be adopted to identify individuals at risk of osteoporotic fracture who will benefit from DEXA examination and preventive therapy [22]. Several case-finding strategies based on clinical risk factors have been developed, such as the Garvan Fracture Risk Calculator, the Women's Health Initiative algorithm, Osteoporosis Self-Assessment tool (OST), Simple Calculated Osteoporosis Risk Estimation (SCORE), Fracture Risk Score, Fracture Risk Calculator, Osteoporosis Risk Assessment Instrument (ORAI), Fracture and Osteoporosis Index of Risk (OSIRIS), and Fracture Risk Assessment Tool (FRAX) [1]. The latter has been used extensively in clinical guidelines and research [22]. FRAX calculates the 10-year probability of major osteoporotic fractures (MOF), defined as hip, wrist, humerus, and spine fractures [36]. The calculation is based on well-established clinical risk factors such as age, sex, body mass index (BMI), prior osteoporotic fracture, family history of hip fracture, current smoking, ever long-term use of oral glucocorticoids, rheumatoid arthritis, other causes of secondary osteoporosis, alcohol consumption of 3 or more units daily [36]. FRAX® can be combined with BMD or T-score to predict the probability of osteoporotic fracture or used alone for case-finding [37]. Different countries use various strategies to identify individuals benefitting from the DEXA examination. In Norway, DEXA should be offered to persons with one or more risk factors [29] or those who score more than 15 % in FRAX® calculation

without BMD [38], while those scoring ≥ 20% should be offered treatment without the BMD test [39, 40]. However, this strategy does not seem well-functioning: only 25% of women and 17 % of men at high risk of osteoporotic fracture (FRAX score greater than 20%) receive anti-osteoporotic treatment [39].

1.5 Bone anatomy and metabolism

The structural unit of bone is called "osteon," or a Haversian system [41]. It has a cylindrical structure with a central canal inside. Concentric lamellae form the walls of the osteon and contain blood vessels and nerves. Cells, called osteocytes, lay between lamellae (Fig.1).

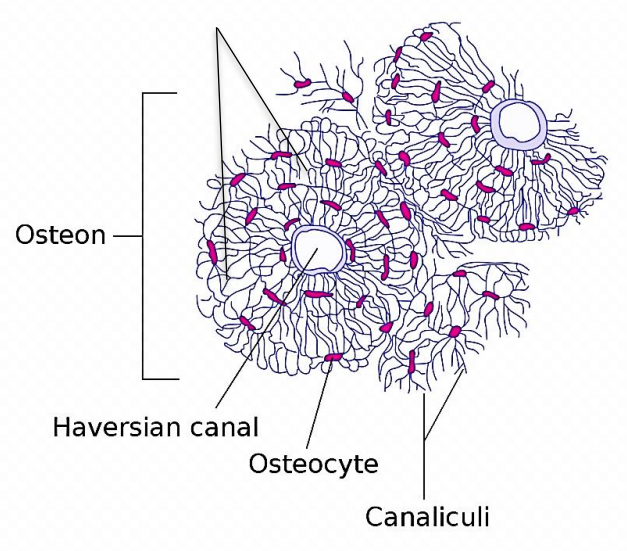


Figure 1 Anatomy of osteon. The image is available from: https://en.wikipedia.org/wiki/Osteon#/media/File:Transverse_section_of_bone_en.svg

All bone tissues in the human body are architecturally categorized into cortical (compact) and trabecular (spongeous) bones. Cortical bone comprises 80% of the total bone in the human body, while trabecular only 20% [41]. Trabecular bone is highly porous (75-95% porosity)

and consists of plates and rods forming trabeculae and surrounding spaces filled with bone marrow. Cortical bone porosity is usually less than 5%. However, healthy adults tend to have thinner cortices and higher cortical porosity during aging [42].

Bone is a dynamic tissue: bone structure undergoes constant remodeling (resorption and formation) throughout life. This process prevents microdamage accumulation in bone tissue [42]. Osteoclasts are the cells responsible for bone resorption, while osteoblasts are responsible for bone formation. During childhood and adolescence, bone formation is greater than bone resorption. These two processes are balanced in young adults, while in older adults, specifically postmenopausal women, bone resorption exceeds bone formation. Besides longitudinal growth in childhood and adolescence, bone also grows radially throughout life: on the periosteal (outer) surface of cortical bone, bone formation occurs faster than resorption, while on the endosteal (inner) surface, the opposite process is observed [42]. Trabecular bone has a ten times larger surface-to-volume ratio and remodels much faster than cortical bone [41]. In the mandible, bone turnover in the alveolar bone is twice as high as in the mandibular body and 3-5 times faster than in the mandibular cortex [43].

Sex hormones define bone growth, bone formation, and resorption to a great extent. In the prepubertal period, bone formation on the outer cortical bone layer increases bone width in both boys and girls [44]. When boys enter puberty, periosteal growth continues in all bones, while endosteal growth occurs slower due to androgens. The opposite process occurs in girls: periosteal bone formation is inhibited, while estrogens stimulate endosteal growth. The cortical thickness is similar in mature young men and women [45]. Periosteal bone formation continues with older age but at a slower rate and has differences in men and women. Bone resorption at the endosteal surface of bones exceeds bone formation similarly in men and women. However, at the outer surface, the bone formation process prevails at a greater rate in

men than in women. The trabecular bone loss also differs by sex: trabecular plates tend to erode and disappear in females while they are just thinning in males. The difference in bone remodeling is conditioned by sex hormones such as estrogens and androgens and partly explains why fragility fractures occur more frequently in females than males [44].

Bone remodeling is affected by mechanical strains or their absence [46]. In their experimental research with mice, Miller et al. showed that the cortical thickness of the tibia on a cross-section changes in response to loading magnitude [47]. A study showed thinning of cortical bone and more intensive loss of trabecular bone at the hip in astronauts 4-6 months of spaceflights [48]. However, the response to mechanical strains is more pronounced in males than females [45].

1.6 Historical review of research on oral bone loss and osteoporosis

In 1960, Groen et al. hypothesized that alveolar bone loss and presenile osteoporosis were related [49]. Hildebolt, in his review article, refers to the earliest histomorphometry studies that emerged in the 1970s and used specimens from cadavers and microradiographs to analyze links between oral and skeletal bone loss [50]. Those studies suggested that secondary hyperparathyroidism, which was known to affect bone metabolism and cause osteoporosis, also influenced the jawbones to a great extent, specifically alveolar processes and supportive tooth structures [50-52]. Subsequent cadaver studies found an increased cortical porosity of the mandible with age and faster bone resorption in the alveolar process than in the mandibular body. They suggested that the alveolar process, trabecular part, and buccal cortex of the mandibular body are more affected by local factors such as mechanical strains and inflammation in supportive tooth structures, i.e., periodontal disease [50, 53, 54]. Then, the trabecular and cortical bone below, anterior mental foramen, and maxillary bone

were investigated in further studies in order to find a stable place to explore age and sex variations in jawbone densities. It turned out that the anterior mandible and maxillary trabecular bone's trabecular bone considerably varied between individuals and thus was considered of limited value [50, 55, 56]. After introducing single and dual photon absorptiometry in clinical practice for measuring bone mineral content (BMC), some studies investigated the relationships between missing teeth or alveolar ridge resorption and BMC in the skeleton in vivo, as pointed out in Hildebolt's review [50]. Researchers found low BMC in females and males with severe alveolar ridge resorption and in edentulous males, referred to as vestibuloplasty [50, 57-59]. At the same time in the 1980s, Bras et al. found that the cortex at the gonial region tends to be thinner at the age of over 60 in men and specifically in women; thus, the cortical thickness of less than 1 mm was considered a useful indicator of metabolic bone disease [60, 61]. Subsequent studies also found that loss of densities in mandibles and radii was more pronounced in females than in males [50, 62, 63]. In 1987, DEXA scanners became available in clinical practice to measure skeletal bone mineral density and diagnose osteoporosis [64]. Jawbone measurements on dental radiographs were validated using DEXA of cadaver mandibles. In an in-vitro study, researchers found that dental bitewing radiographs reflect changes in bone mineral density measured by DEXA when gradually removing layers of bone tissue on the cadaver mandibles. Thus, bitewings were suggested for osteoporosis patient identification [65]. Subsequent studies investigated the link between jawbone measurements on different dental radiograph types and DEXA measurements at various skeletal sites [50].

1.7 Radiomorphometric indices

Several methods have been developed to quantify jawbone mass on dental radiographs for predicting osteoporosis. Some of these methods focused on the morphology of the inferior mandibular cortex on dental panoramic radiographs: Mandibular cortical width (MCW) proposed by Ledgerton et al. [66], Panoramic mandibular indexes proposed by Benson et al. [67], and Mandibular cortical index (MCI) proposed by Klemetti et al. [68] and Gonial Index proposed by Bras et al. [61]. Another method, called the visual radiographic index, proposed by Lindh et al., estimated the architecture of the mandibular trabecular bone assessed mainly on intraoral dental radiographs [69]. The most widely used indices in the scientific literature are MCW and MCI, measured on dental panoramic radiographs [70]. According to Ledgerton et al., MCW is measured on dental panoramic radiographs along the line drawn through the middle of the mental foramen and perpendicular to the lower mandibular border (Fig.2) [66]. According to Klemetti et al., MCI classifies the mandibular cortex into dense (C1), moderately eroded (C2), or severely eroded (C3) (Fig.3) [68].

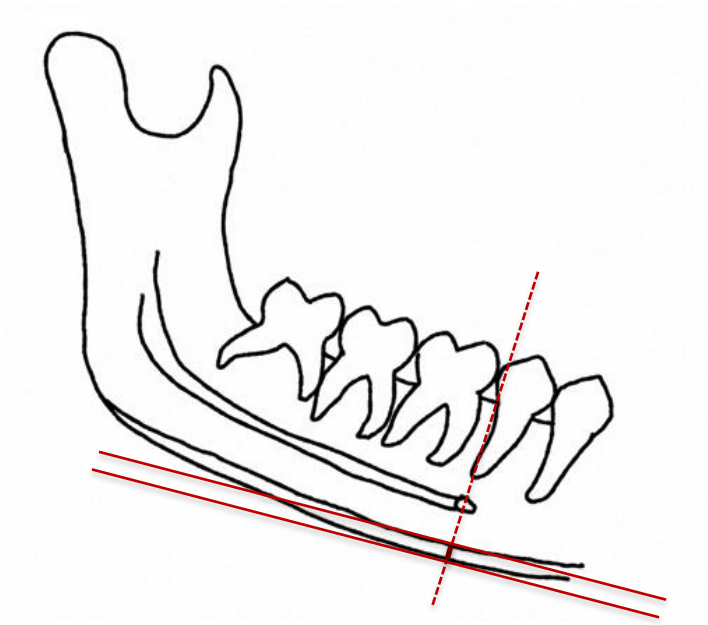


Figure 2 Mandibular cortical width measurement

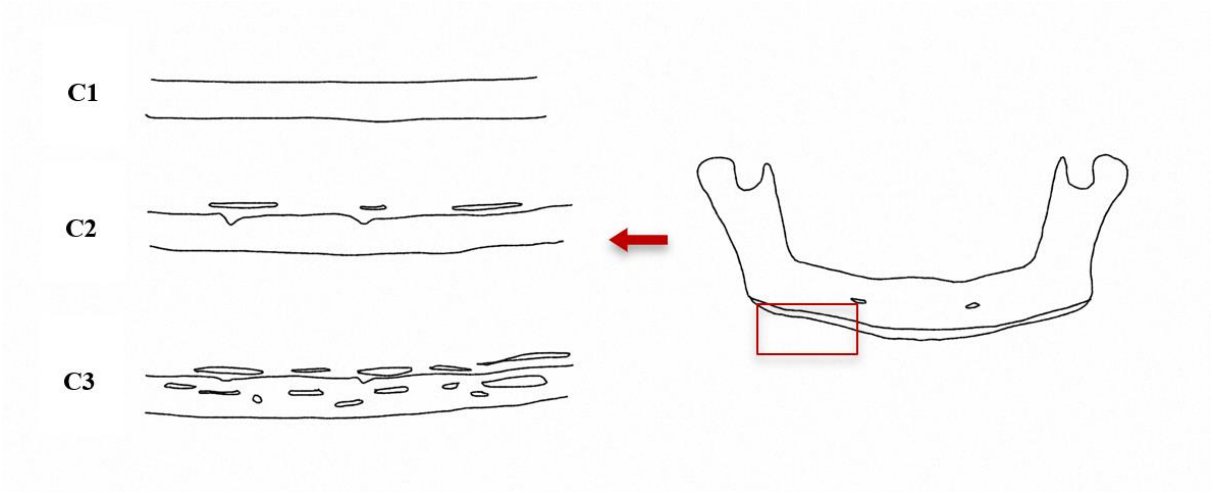


Figure 3 Mandibular cortical index assessment

1.8 Relationship between radiomorphometric indices, skeletal BMD, and other factors

Previous studies established a significant weak to moderate correlation between radiomorphometric indices and skeletal bone mineral density at the lumbar spine [71-73], femoral necks [72, 74], forearm[72], and calcaneus [75]. These findings suggested that skeletal bone mineral density status is related to mandibular cortical morphology. Several studies investigated factors affecting MCW and MCI. Some of them showed that older people tended to have thinner mandibular cortexes as measured by MCW, and females experienced a more pronounced thinning than males [76-78]. One longitudinal study explored the relationship between age and cortical erosion (MCI) and found that the cortex became eroded at least by one category in 68.9% of study participants independently on their sex in an average of almost two decades [79]. Several studies examined the impact of missing teeth on cortical thinning and erosion [75, 78, 80-82]. Some found an independent impact of missing teeth in both males and females [75, 81, 82], while one study found an association only for females [80]. Periodontal disease is another factor that can affect jawbone quality through

different mechanisms, such as inflammatory mediators activating bone resorption or tooth loss, resulting in reduced mechanical strains and subsequent bone loss [83-85]. Studies exploring the link between osteoporosis and periodontitis exist, but their results are still controversial, although most of them found an association between the diseases [86]. Few studies compared radiomorphometric indices in females with periodontitis and healthy females. One found a more eroded cortex in females with periodontitis, while another did not [87, 88]. However, many of the mentioned studies lacked adjustment for confounders and did not include all possible predictors listed above [66, 75, 79, 87, 88]. Thus, it is unclear which factors predict MCW and MCI and how much they contribute to mandibular cortical morphology.

A number of studies investigated the diagnostic efficacy of radiomorphometric indices on various thresholds for identifying individuals with osteopenia and osteoporosis. Calciolari et al., in a systematic review with a meta-analysis, found 34 studies using MCW where thresholds varied from 2.69 mm to 5 mm, sensitivity from 12% to 94%, and specificity from 55% to 100% for osteoporosis or osteopenia diagnosis [89]. The pooled sensitivity and specificity for MCW were 43% and 90%, respectively, indicating that MCW could better identify healthy patients [89]. The same systematic review found 27 studies analyzing MCI. The presence of any cortical erosion (C2 or C3 categories) showed sensitivity ranging from 59% to 100% and specificity ranging from 45 to 91% for osteoporosis or osteopenia diagnosis in the included studies. The pooled sensitivity and specificity were 80% and 64%, respectively [89]. Kinalski et al., in a more recent systematic review, found similar results for MCW and MCI [90]. However, they calculated the pooled diagnostic efficacy of these two indices for osteoporosis diagnosis separately from osteopenia: MCW presented a sensitivity ranging from 10% to 90% and specificity ranging from 34% to 100%; The pooled sensitivity

was 62%, while the specificity was 77% [90]. The sensitivity of MCI varied from 16% to 87%, and specificity varied from 64% to 100% in the included studies. The pooled sensitivity and specificity values were 35% and 88%, respectively [90]. Two systematic reviews concluded that MCW and MCI could be used for low bone density diagnostics [89, 90]. However, low bone density or osteopenia, primarily explored in the previous studies, is not considered a disease and, therefore, is not of clinical interest [22, 89]. Among the other major limitations, both systematic reviews revealed in many studies a high risk of bias in the participant's selection domain, i.e., lack of random sampling. Most studies utilized a case-control design, convenience, or consecutive sampling procedure that reduces their external validity [89, 90]. Sample size problems were also inherent to the studies on the diagnostic efficacy of radiomorphometric indices. Some studies included 100 or fewer participants [72, 74, 91, 92]. Thus, the quality of existing evidence is generally suboptimal. Kinalski et al. recommended conducting further studies with a random selection of study participants [90].

One of the most prominent studies in this area was the OTEODENT project, a multi-center study exploring various aspects of radiomorphometric indices in a large study sample. They recruited 670 females in four European centers in Sweden, the United Kingdom, Greece, and Belgium. The researchers found that MCW performed better than MCI and that the optimal MCW threshold was 3 mm, i.e., those with less than 3 mm were considered at risk of osteoporosis [93]. They also concluded that the combination of MCW and a risk factor-based tool for osteoporosis assessment, OSIRIS, can be helpful in settings where higher specificity is preferred [94]. However, radiomorphometric indices should only be recommended as an additional tool that utilizes dental radiographs taken for other indications, not for osteoporosis diagnosis [93]. The researchers in the OSTEODENT project also found an association between osteoporosis and tooth loss independent of age and smoking status [95].

Taguchi et al. are the other researchers who devoted many articles to exploring the links between osteoporosis and changes in dental radiographs [71, 73, 91, 96, 97] and even wrote clinical guidelines for osteoporosis screening in Japanese dental clinics [98]. They found that those with MCW in the lowermost quartile had 5.43 higher odds of having low vertebral BMD or osteoporosis than the uppermost quartile[71]. At the same time, those with severely eroded cortex had 17.73 times higher odds of osteoporosis than participants with dense cortex [71]. They also found that MCW and MCI could distinguish between healthy and low BMD individuals among younger postmenopausal women [96].

1.9 Knowledge gap

The combination of clinical risk factors for osteoporosis with radiomorphometric indices has not been extensively studied in the abovementioned studies, while the systematic reviews recommended further studies doing so [89, 90]. The most comprehensive study that did it was the OSTEODENT project, in which the researchers combined MCW and OSIRIS [94]. However, FRAX is a more widely used tool recommended by the European guidance, the International Osteoporosis Foundation, and the American Task Force for osteoporosis case finding [22, 99, 100]. Thus, exploring the combination of FRAX with radiomorphometric indices is reasonable. FRAX has been used in previous studies on radiomorphometric indices, but mainly as a reference standard or an outcome, not as a case-finding tool [101, 102]. One study examined combining radiomorphometric indices with FRAX for osteoporotic fracture prediction and found a better predictive ability for FRAX and MCI than for either used alone [103]. However, this study used self-reported fractures; no BMD measurements were available [103].

The validity of radiomorphometric indices for male osteoporosis screening has also been poorly studied. The prevalence of osteoporosis in males is lower than in females [14]. A study investigating the diagnostic efficacy of radiomorphometric indices would require a representative sample of males with the whole spectrum of BMDs; such a study would need to recruit more than one thousand males, which can be challenging. Nevertheless, a smaller sample would be enough to assess the strength of the relationship between BMD and mandibular cortical indices. Very few studies investigated radiomorphometric indices in males [104-106]. The most extensive study was conducted by Leite et al. and included 127 males. They found a sensitivity of 65%, specificity of 90%, and area under the curve (AUC) of 0.77 for MCW in males. For MCI (C3), sensitivity and specificity were 62% and 100%, respectively. However, the study recruited non-random participants: the prevalence of osteoporosis was 50% [106]. Thus, the more severe osteoporotic cases were likely to be recruited, making the study results biased. Two other studies had fewer participants (40 and 49) and a lack of adjustments for confounders such as age [104, 105]. Thus, no conclusions about whether radiomorphometric indices are valid for osteoporosis screening in men can be drawn.

1.10 Manual and computer-aided radiomorphometric indices

All the abovementioned studies used ordinary or digital calipers to measure MCW manually, and all MCI assessments were visual. Thus, all radiomorphometric measurements were observer-dependent. The main concern of those studies, also mentioned in the systematic review by Calciolari et al. [89], was low intra- and interobserver agreement. Taguchi explored the intraobserver agreement between 60 observers from different countries assessing MCI. More than 60 % showed weighted kappa values of more than 0.6, and only three observers

showed a fair or poor agreement. The researchers concluded that it was a sufficient agreement [107]. Devlin and Horner also explored variations in MCW and MCI assessments made by nine general dental practitioners compared to dental radiologists. Unlike Taguchi et al., they found a substantial variation in MCW measurements between nine general dental practitioners and dental radiology specialists. The mean differences in measurements ranged from 0.67 mm to 1.64 mm, while their limits of agreement were several millimeters in range. Weighted kappa for MCI ranged from very good to poor for different observers [108]. The OSTEODENT project also explored the reliability of radiomorphometric indices and concluded that the agreement was suboptimal and that the possibility of automated detection had to be explored [94, 109].

The first known study trying to develop an automatic measurement of MCW was done by Arifin et al. They used an image processing algorithm, but users must indicate the region of interest, identify mental foramen, and follow the algorithm. Thus, it was a semi-automatic computer-aided method. One hundred postmenopausal women were involved; a high sensitivity of 88% was found, while specificity was 57.5% [110]. Later, they proceeded with similar work and developed other algorithms, such as Fuzzy Neural Network combining MCW and MCI [111]. Allen et al. developed a computer-aided system based on active shape models in a British sample [112]. Then, a group of authors involved in the OSTEODENT tested the abovementioned algorithm in their study sample [113, 114] and even combined it with the risk factor tool OSIRIS [115]. Nakamoto et al. developed a computer-based MCW assessment system based on skeletonization and subsequent analyses of the grayscale values of the mandibular cortex [116]. Recently, studies using artificial intelligence and deep learning have emerged [117, 118]. Some studies concluded that computer-aided measurements performed as well as experts or risk factor-based tools [110, 113], while others

found a superior activity compared to manual measurements [117-119]. Similar limitations, specifically low sample size, were present in the studies utilizing computer-aided measurements.

2 Objectives

This thesis aims to evaluate the utility of panoramic radiomorphometric indices (MCW and MCI) for finding Norwegian females and males aged 40 and older at high risk of osteoporosis who can benefit from further DEXA examination.

Specific objectives were:

Paper I

1. To evaluate the diagnostic efficacy of radiomorphometric indices in adult males and females living in Tromsø
2. To evaluate the diagnostic efficacy of radiomorphometric indices applied to a population of older adults (≥65)
3. To evaluate the diagnostic efficacy of radiomorphometric indices combined with a clinical risk-factor-based tool, FRAX.

Paper II

1. Evaluate relationships between general factors such as T-score, age, and menopausal status (for females), local factors such as the number of remaining teeth and periodontal status, and the morphology of the mandibular cortex.

Paper III

1. To explore the feasibility of detecting mental foramen on dental panoramic radiographs with pre-trained object detection models and to investigate the possibility of developing an automatic algorithm measuring MCW.

3 Materials and methods

This study is based on data from an ongoing population-based study called The Tromsø Study and carried out as repeated cross-sectional surveys in Tromsø, Norway (Fig.4) [120]. The data from the seventh survey of the Tromsø study carried out in 2015-2016 (Tromsø7) have been used in this doctoral thesis [121].

Figure 4 Tromsø, Norway. The image available from: https://www.kartverket.no/

3.1 Study population

All people aged 40 years and older (n=32,591) living in Tromsø were invited to participate in Tromsø7. Twenty-one thousand and eighty-three persons consented to participate, yielding a response rate of 65% [121]. They completed extensive questionnaires covering socio-demographic characteristics and various health-related topics. An extensive clinical examination took place during the first visit. Anthropometric measurements, biological

samples, cardiovascular parameters, and dental clinical examination were conducted [121]. Three thousand nine hundred fifty-one study participants out of 21,083 attending the first visit underwent dental panoramic radiograph examination (Fig.5). Eight thousand three hundred sixty randomly selected from 21,083 study participants attended the second clinical visit. They underwent other clinical tests, including cognitive, eye, lung, heart, and body measurements. Among those participants, 3,600 underwent a DEXA examination (Fig.5) [121].

To answer the research questions in this thesis, the participants who underwent both dental panoramic radiography and DEXA were eligible for **Paper I** and **II**. All participants who received dental panoramic radiography were eligible for **Paper III**. The flow of the study participants in **Papers I, II, and III** is described in detail in Figure 5. The current project, including three studies, was initiated in 2019.

3.2 DEXA

DEXA scan at both hips and femoral necks was conducted by a Lunar Prodigy device (G.E. Healthcare Lunar, Madison, WI, USA). Trained technicians inspected the DEXA scans according to the manufacturer's instructions. The equipment underwent daily calibrations using a standard phantom. Areal bone mineral density (g/cm^2) and T-score were registered. This study used a standard Lunar reference population drawn from the NHANES III sample to calculate a T-score for each participant. The diagnosis of osteoporosis was defined according to the European guidelines for diagnosing and managing osteoporosis (T-score $\leq$ -2.5 SD) [22]. T-scores at the left and right femoral necks were used to identify osteoporosis. If at least one side met the criteria for the disease, the participant was considered an osteoporosis case in **Paper I**. The lowest T-score on either side was used as a continuous

predictor variable in **Paper II**. The observers were blinded to DEXA measurements during the assessment of MCW and MCI.

3.3 Dental panoramic radiographs

Dental panoramic radiographs (DPRs) were taken from the participants using Planmeca-ProMax-s3-2d-dimax-4-pano (Planmeca Oy, Helsinki, Finland). Exposure parameters were set to 68 kV, 8 mA, and 16 s but could be adjusted to the patient's size. The receptor was a charge-coupled device having a pixel size of 99x99 μm^2 with a digital 16-bit grey-level output. The images were stored in the Tag Image File Format (TIFF) with a spatial resolution of 101.1 pixels per cm, equivalent to 5-line pairs per millimeter or 257 dots per inch. DPRs were 2821×1376 pixels.

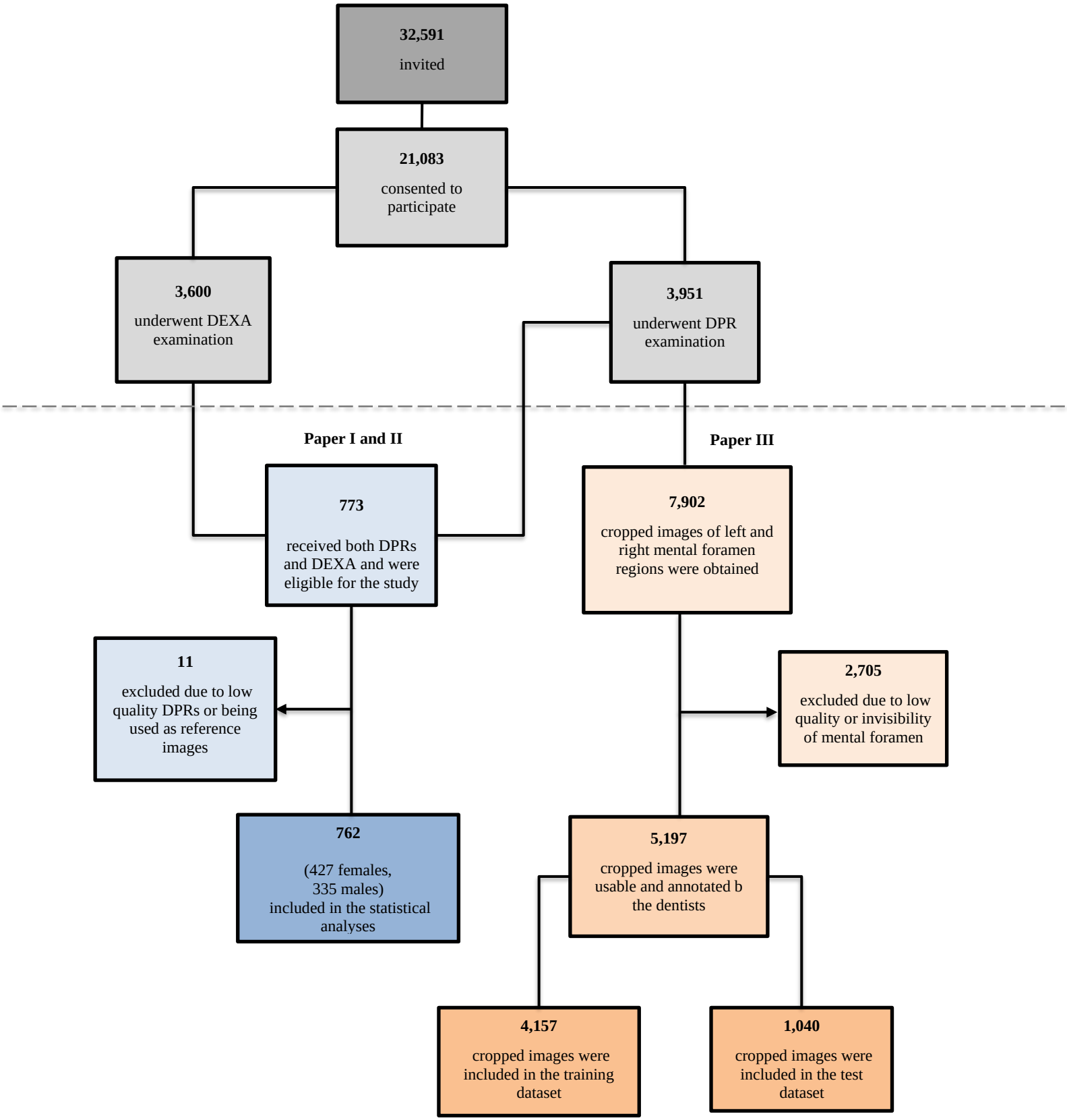


Figure 5 Selection of study participants

3.4 MCW measurements

MCW was measured on 762 DPRs on both sides of the mandible using ImageJ, 1.8.0172 software (U.S. National Institutes of Health, Bethesda, Maryland, USA) [122]. A line was drawn through the middle of the mental foramen and perpendicular to the tangent to the lower mandibular border, as proposed by the original method [66]. If the mental foramen was not distinguishable, the line was drawn between the apices of premolars. If premolars and mental foramen were absent on a DPR, the line was drawn approximately in the premolar region. MCW was measured in mm along this line. Then, the image was processed in ImageJ with the following steps:

1. The image was normalized to get the full spectrum of images' gray value intensities.
2. The region of interest (ROI) was manually selected in the mental foramen area, including the mandibular cortex (Fig.6, A). The selected ROI was subsequently re-normalized.
3. The line perpendicular to the lower mandible edge was drawn (Fig.6, B).
4. A Gaussian blur filter was applied.
5. A grayscale plot was made where the lower values represented the darkest area of the image (Fig.6, C).
6. Turning points in the grayscale plot represented the upper border of the cortex (A) and the lower border of the cortex (B) (Fig.6, C). The distance A_1B_1 corresponded to the thickness of the cortex.

Since reference no reference object were used to control for magnification, MCW measurements were adjusted for a standard magnification factor of 20% indicated in the panoramic unit manual. Overexposed, underexposed, or unsharp dental panoramic

radiographs were excluded (Fig.5). MCW measurements were used as an index test in **Paper I** and as an outcome variable in **Paper II.**

3.5 MCI assessment

DPRs were first normalized, then the cortical shape was evaluated distally from the mental foramen on both sides of the mandible using the MCI proposed by Klemetti et al. [68]. The mandibular cortical bone was classified according to the original method as C1 – even and endosteal scarp margin on both sides (Fig.7, A); C2 – some endosteal cortical residues and semilunar defects on one or both sides (Fig.7 B); C3 – heavy endosteal cortical residues, the cortical bone is porous (Fig.7, C) [68]. In addition, three reference images from the study participants with known T-scores were used as examples where C1 corresponded to normal T-score, C2 corresponded to osteopenia, and C3 to osteoporosis (Fig.7). MCI measurements were used as an index test in **Paper I** and as an outcome variable in **Paper II.**

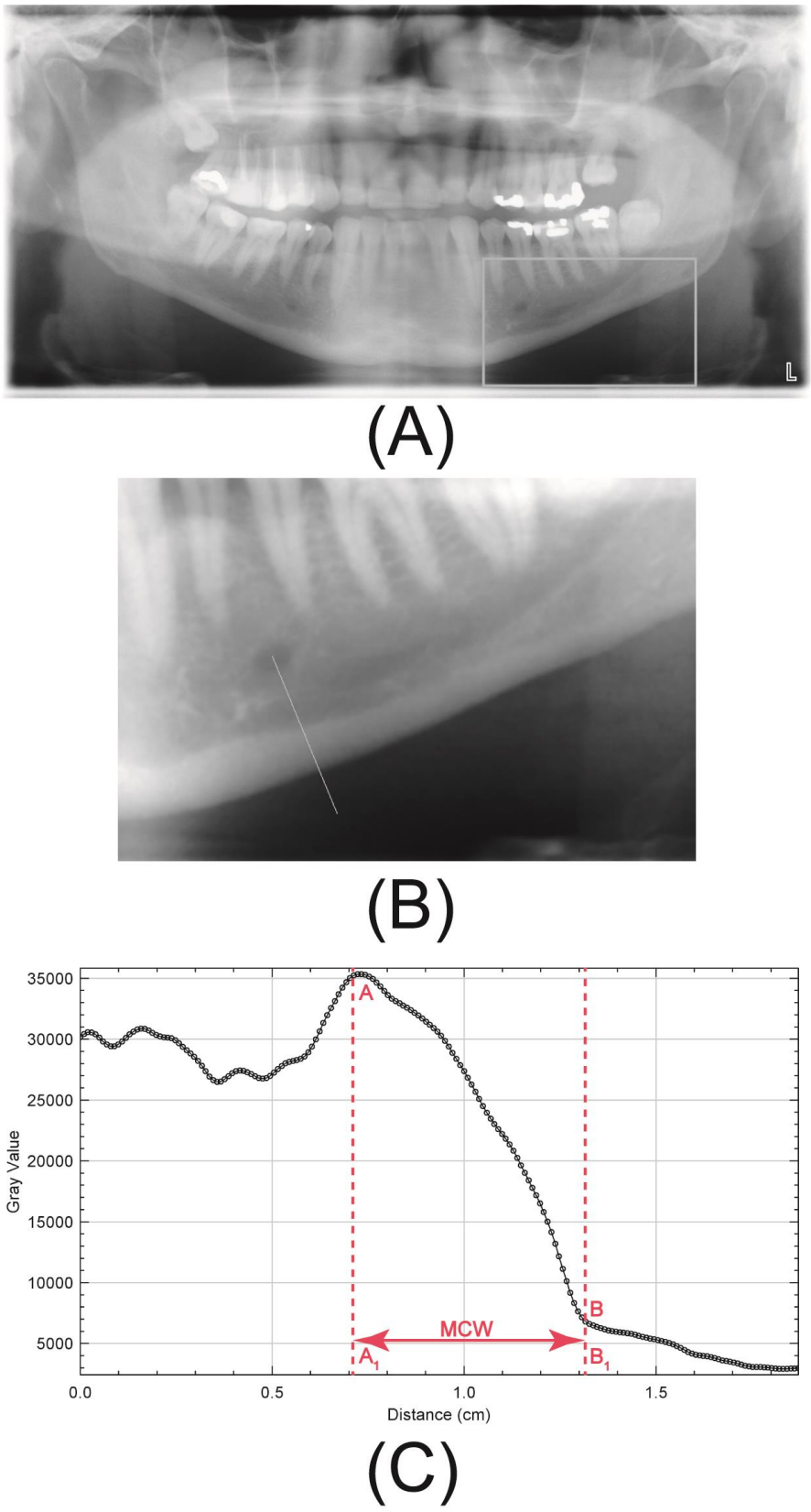


Figure 6 Mandibular cortical width measurement using ImageJ

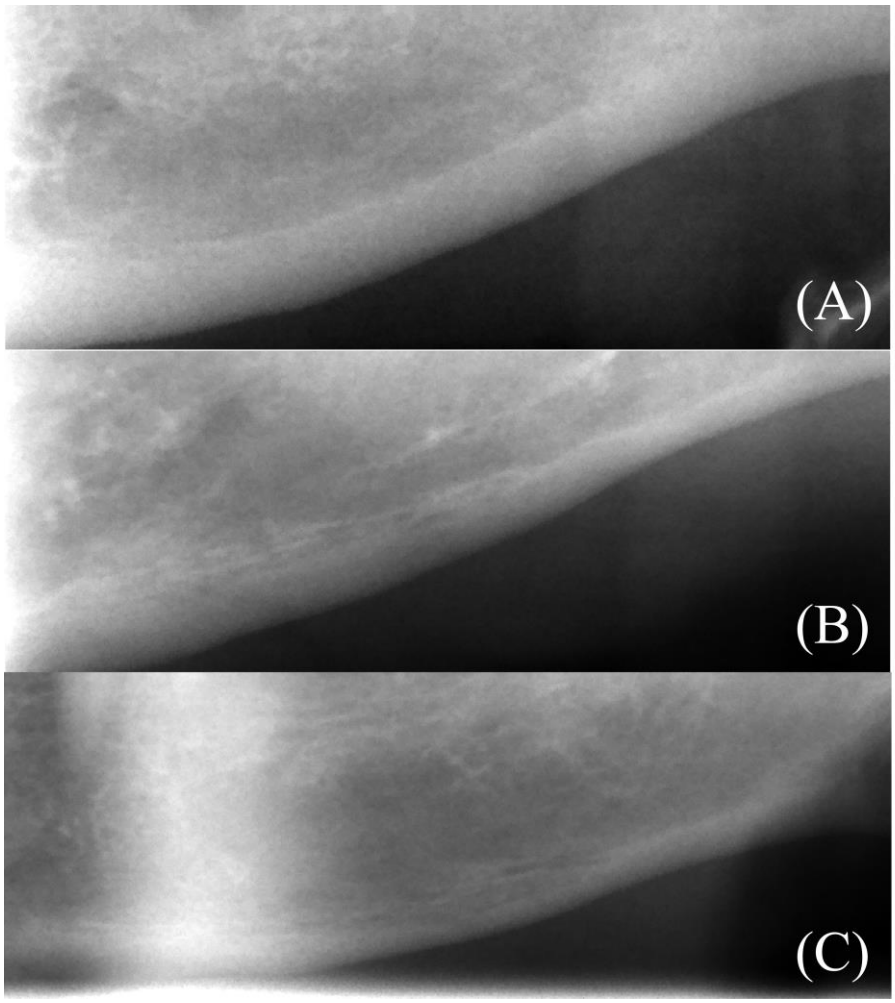


Figure 7 Reference images for mandibular cortical shape assessment

3.6 Computer-aided mental foramen localization and MCW measurement algorithm

In **paper III,** a supervised machine learning method was used to make an algorithm locating mental foramen (MF), and it was further combined with an automated measurement of MCW for osteoporosis prescreening. Machine learning is a subfield of artificial intelligence aimed at solving tasks without pre-defined rules and instead learning those rules from data. Supervised machine learning is one of the machine learning methods commonly used for object

localization in medical imaging [123]. “Supervised” means that data are labeled beforehand, i.e., the algorithm is provided with a desired output.

All 3,951 DPRs were used. Two ROIs, i.e., the left and right sides of the mandible containing MF and cortex, were automatically cropped, yielding 7,902 crops in total. The size of each crop was 300 × 600 (height ×width) pixels. Two thousand seven hundred-five crops were excluded due to not capturing the jaw cortex and MF, being distorted, overlapping anatomical structures in ROI, or experts being unable to label mental foramen. A total of 5,197 crops were used in the study; 4,157 were allocated in the training dataset, and 1,040 in the test dataset. Two dental experts annotated all crops using VIA, the Annotation Software for Images, Audio, and Video, and drew a bounding box around MF [124]. They divided the workload to save time. However, 706 crops were annotated by both dentists to assess their agreement. The dentists also picked 100 “easy images” where MF was visible and another 101 “complex images” where mental foramen was challenging to distinguish. It was done to test further how the best model will work under different circumstances. Four existing object detectors pre-trained on the COCO data set [125] were used:

1. Faster R-CNN with ResNet50 [126]
2. CenterNet with HourGlass104 [127]
3. EfficientDet-D0 [128]
4. RetinaNet with ResNet50 [129]

The object detectors were fine-tuned using the TensorFlow framework [130]. “Fine-tuning” refers to further training existing object-detection models with similar tasks but on different data. Two different experiment setups were tested for each object detector. After fine-tuning the training dataset, the object detectors were evaluated on the test dataset. The model

performing best at mental foramen localization was combined with an algorithm automatically measuring MCW based on classical image processing, involving filtering thresholding and grayscale analyses. The details on the image processing algorithm for MCW measurement are available in the publication [131].

3.7 Other risk factors

All participants reported age and sex in the questionnaires. Age was combined with MCW or MCI and used as a part of the index test in **Paper I** and as a predictor variable in **Paper II**. In **Papers I and II,** all analyses were carried out separately for males and females.

Height and weight were measured during the first visit. The BMI of an individual participant was calculated by the weight in kilograms divided by the square of the height in meters. The study participants reported smoking, alcohol use frequency and amount, rheumatoid arthritis, and medicine use in the questionnaires. The same study participants reported information on previous osteoporotic fractures and osteoporosis in parents in the previous survey (Tromsø6) but not in Tromsø7. Therefore, we retrieved this information from Tromsø6. The information on osteoporosis in parents was used as a substitute for a family history of hip fractures. FRAX score was calculated using the online tool based on the following risk factors: age, sex, BMI, family history of hip fractures, previous osteoporotic fracture, current smoking, glucocorticoid use, rheumatoid arthritis, and alcohol use three or more units a day [37]. FRAX was combined with MCW or MCI and used as index tests in **Paper I**.

Menopausal status was reported in questionnaires in Tromsø7 and used as a predictor variable in **Paper II.** Missing teeth and periodontal status were recorded during the dental examination on the first visit**.** Periodontal probing depth (PPD) and bleeding on probing (BOP) were

recorded. PPD was measured by the World Health Organisation periodontal probe (LM55B). Radiographic marginal bone level (RBL) was measured on DPRs, as proposed by Holde et al. [132]. Periodontitis was defined using the American Academy of Periodontology classification launched in 2017 [133]. If RBL was detected in two or more adjacent teeth, it was considered a periodontitis case. Stage I was defined if RBL<15% and PPD ≤4 mm, stage II if RBL was 15-33% and PPD ≤5 mm, or RBL<15% and PPD 5 mm, and stage III-IV if RBL>33%, or RBL<33% and PPD ≥6 mm. Stages III and IV were collapsed into one category. Periodontal stability was defined in the case of RBL detected on two or more non-adjacent teeth, but PPD ≤ 3mm. Periodontium was considered healthy if BOP < 10%, and gingivitis was defined if BOP ≥ 10% [134].

3.8 Inter- and intraobserver agreement of MCW and MCI

Two observers, a general dental practitioner and an oral radiologist, had a one-hour training session in measuring MCW and MCI on 20 DPRs. After that, they independently assessed MCW and MCI on another 50 DPRs randomly selected from the dataset of 3,951 images. Then, one of the observers repeated measurements in one month using the same 50 images plus another randomly selected 50 DPRs from the same dataset. The sample size for inter- and intraobserver agreement calculation was arbitrarily chosen for feasibility reasons.

Since MCW was a continuous variable, Bland-Altman plots were used to analyze inter- and intraobserver agreements. A Bland-Altman plot shows the mean difference in two MCW measurements and its limits of agreement with 95% confidence intervals [135]. Linear regression analysis between the difference in MCW measurements and mean MCW between two observers was used to check a systematic error in MCW measurements. Clinically acceptable limits of agreement were defined as ±0.8 mm based on the difference in average

MCW between osteoporotic and healthy females. Since MCI had three categories (C1, C2, and C3), weighted kappa was used to calculate inter-and intraobserver agreement. Landis and Koch's criteria for agreement were used to interpret the results [136]. Inter- and intraobserver agreements were reported in **Paper I.**

Bland-Altman plots for MCW measurements are presented in Figures 8 and 9. As linear regression analysis showed, there were no systematic errors, i.e., changes in the mean difference in MCW measurements between two observers with increasing mean MCW. One sample t-test showed that the mean difference in MCW measurements was equal to 0 in inter- and intraobserver agreement analyses (p=0.3 and p=0.45). However, the confidence intervals for the limits of agreement exceeded clinically acceptable ±0.8mm in both analyses (Fig.8, 9).

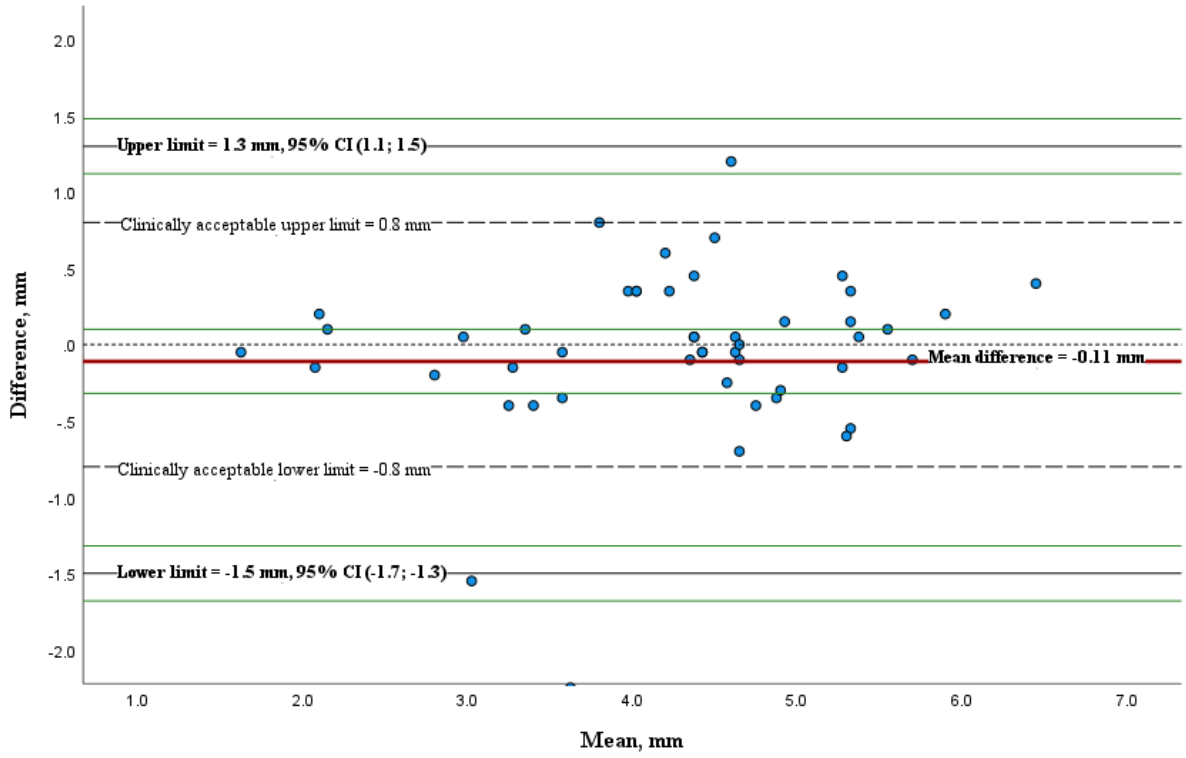


Figure 8 Bland-Altman plot for interobserver reliability

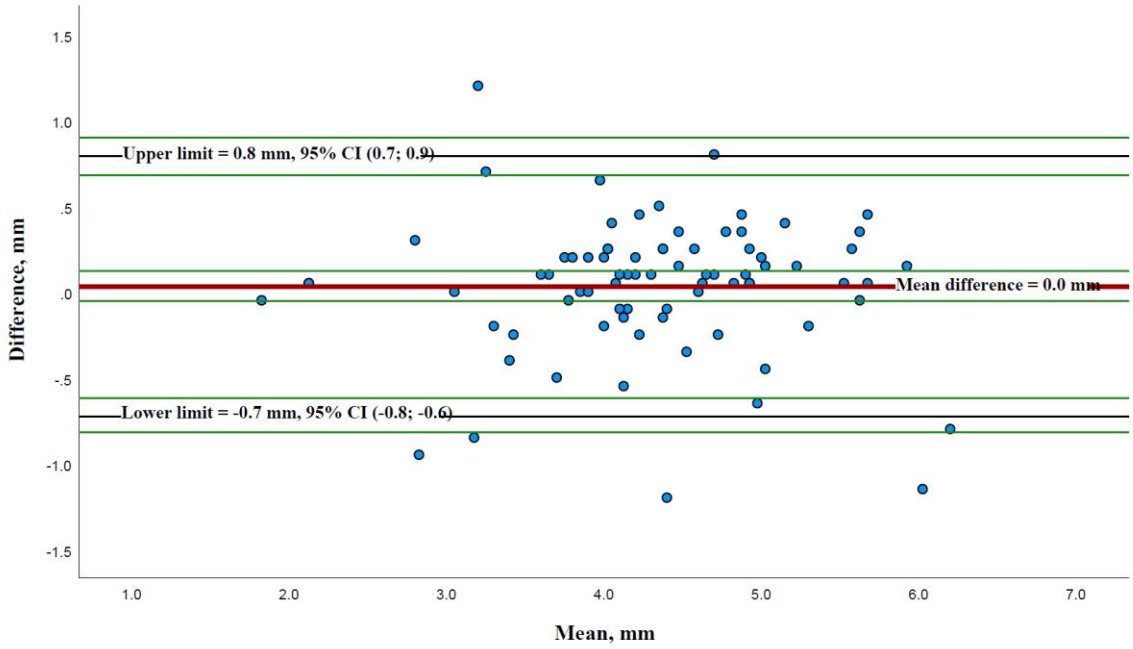


Figure 9 Bland-Altman plot for intraobserver reliability

Inter- and intraobserver agreements for MCI measured by weighted kappa were 0.47 [95% CI = (0.30-0.65) ($P < .05$)] and 0.72 [95% CI = 0.59-0.85 ($P < .05$)], indicating moderate and substantial agreement, respectively [136].

3.9 Statistical analysis, Paper I

Diagnostic efficacy parameters, *i.e.*, sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (+LR), and negative likelihood ratio (-L.R.), were calculated. Sensitivity is the proportion of true positive test results among all participants with osteoporosis. Specificity is the proportion of true negative test results among all healthy participants. Accuracy is the proportion of correct test results among all test results. PPV is the proportion of participants with osteoporosis among all participants with positive test results. NPV is the proportion of healthy participants among all

participants with a negative test result. Positive LR is the probability of getting a positive test result in a participant with osteoporosis divided by the probability of a positive test result in a healthy participant. Negative LR is the probability of getting a negative test result in a participant with osteoporosis divided by the probability of getting a negative test result in a healthy participant.

The diagnostic threshold for MCW was set to be 3 mm. Two diagnostic thresholds were set for MCI: first, mild or severe erosion, C2-C3, and second, severe erosion, C3. Diagnostic efficacy was calculated for the index tests and their combinations: 1) MCW ≤ 3mm, or MCI C2-C3 or C3; 2) MCW ≤ 3mm, or MCI C2-C3 or C3 and age ≥65 years; 3) MCW ≤ 3mm, or MCI C2-C3 or C3 and FRAX-score > 15%. ROC curve analysis was performed only for MCW, not combined with other parameters. Sensitivity was plotted against the false positive rate at different MCW thresholds, and the area under the curve was calculated. The analyses for males and females were separated.

SPSS version 26.0. (IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Armonk, NY: IBM Corp) was used to calculate weighted kappa, Bland-Altman plots with regression analysis, weighted kappa, and ROC curves. MedCalc diagnostic test evaluation calculator, version 20.014 (MedCalc Software, Ostend, Belgium), was used to calculate sensitivity, specificity, probability of a correct test result (accuracy), PPV, NPV, positive likelihood ratio (+LR), negative likelihood ratio (-LR), and their confidence intervals.

3.10 Statistical analysis, Paper II

The distributions of continuous variables were checked for normality using a visual examination of normality plots, skewness, kurtosis, and Q-Q plots. MCW, T-score, and age were normally distributed, while the number of remaining teeth was not. A Pearson’s

correlation coefficient was used to analyze relationships with MCW for all normally distributed variables and MCW. Spearman's correlation coefficient was used for the remaining teeth.

Simple and multiple linear regression analyses were used to analyze the relationship between age, T-score, menopausal status (for females), remaining teeth, and periodontal status with MCW. A hierarchical model with three blocks was used for multiple linear regression. Age was added to the first block, T-score and menopausal status for females were added to the second block, and remaining teeth and periodontal status were added to the third block. MCW, age, T-score, and remaining teeth were continuous variables. The T-score was an ordinary variable with a 0.1 SD increment. Menopausal status was a binary variable while periodontal status was an ordinal variable with the following categories: 1) healthy or mild periodontitis, i.e., healthy periodontium, gingivitis or stage I; 2) moderate periodontitis, Stage II, and 3) severe periodontitis, Stage III-IV. Patients with periodontal stability were excluded. The assumptions of normality, homoscedasticity, and independent errors were checked. Influential cases and collinearity were tested.

Binary and multiple logistic regression analyses were used to analyze relationships of age, T-score, menopausal status (in females), remaining teeth, and periodontal status with MCI. A hierarchical model with the same blocks was used in multiple regression analyses. MCI was used as a binary outcome (C1 vs. C2-C3). Age was used as 10-year age groups: 40-49, 50-59, 60-69, 70-79, and 80+. The oldest participant was 84 years old. T-score and remaining teeth were used as continuous variables. Menopausal status and periodontal status were used as binary and ordinary variables, respectively, as described above. The assumptions of logistic regression were checked: influential cases, multicollinearity, linearity of logit, etc.

SPSS version 26.0. (IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Armonk, NY: IBM Corp) was used.

3.11 Statistical analysis, Paper III

Intersection over union (IoU) is a standard metric allowing the definition of «correct» or «incorrect» object detections and subsequent calculation of the accuracy of an object detector [137]. The method calculates the overlapping area between two bounding boxes; one is produced by an object detector, while another represents «ground truth,» i.e., a reference to compare with. (Fig. 10) [138]. In the current study, the results produced by the object detectors were compared to the dentist's annotations, using IoU at 0.5 and 0.75 thresholds. If the overlapping of bounding boxes produced by dentists and the algorithm was more than 50% or 75%, the results of mental foramen localization were considered «correct.» IoU metric was also utilized to check the agreement between two dental experts since their subjective labels were considered a “ground truth.” Minimum, maximum, and average IoUs, defining the overlap between the two annotations made by the experts, were calculated for 706 images that both experts assessed.

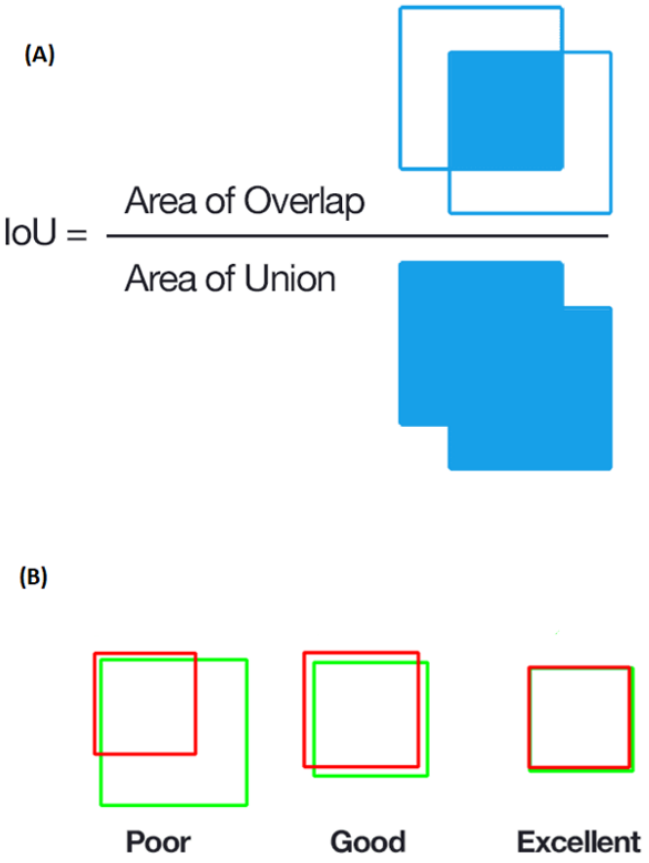


Figure 10 (A) Calculation of the Intersection over Union (IoU), (B) defining a threshold for IoU to consider a detection correct or incorrect. The image was obtained from: https://en.wikipedia.org/wiki/Jaccard_index

The following accuracy metrics are commonly used to assess the performance of an object detector: accuracy, recall, precision, and mean average precision. Accuracy is the total proportion of cases when a model produces correct output. Precision is somewhat similar to PPV and refers to the proportion of positive output that was correct. The recall is the proportion of correct positive outputs among all ground truths, i.e., the metric is similar to sensitivity. One can plot the true positive vs. false positive rate, i.e., the ROC curve, and calculate the AU(AUC is commonly called average or mean average precision [125].

Mean average precision (mAP) was calculated for four object detectors trained in two different setups and tested at both IoU thresholds. Then, the dental experts visually assessed the results produced by the best object detector on 100 “easy images,” and 101 “complex images,” Crosstable and weighted kappa were calculated to present the agreement between dentists' assessments of algorithm output. Then, the best-performing object detector was combined with another algorithm measuring MCW, and its ability to produce output was tested on 100 random images. Further technical information on the model training process, experiment setups, and the mandibular cortical width measurement algorithm is available in the publication [131].

3.12 Summary of tests and variables used

Paper I	Paper II	Paper III
Index tests: 1. MCW, MCI 2. MCW or MCI combined with age ≥ 65 3. MCW or MCI combined with FRAX >15%	**Predictors:** T-score, age, menopausal status (females), remaining teeth, periodontal status	**Tests:** Four pre-trained models were fine-tuned to locate mental foramen: • Faster R-CNN • RetinaNet • CenterNet • EfficientDet-D0
Reference standard: T-score at the femoral neck	**Outcomes:** MCW, MCI	**Reference standards:** Annotations of mental foramen made by experts
Study design: Cross-sectional, diagnostic efficacy	**Study design:** Cross-sectional	**Study design:** Cross-sectional
Statistical analysis: Sensitivity, Specificity, Positive predictive, Negative predictive values, Positive likelihood, Negative likelihood ratio, and receiver operating characteristic curves were calculated	**Statistical analysis:** Linear regression for MCW and logistic regression for MCI	**Statistical analyses:** IoU to define "correct" and "incorrect" MF detections. Mean average precision.

3.13 Funding

The Arctic University of Norway (UiT), Northern Norway Regional Health Authority (Helse Nord RHF), the University Hospital of North Norway (UNN), and different research funds financed the Tromsø study. The Department of Clinical Dentistry, the Faculty of Health Science (UiT), fully financed the current project.

3.14 Ethical considerations

The Tromsø Study was conducted in accordance with the World Medical Association Declaration of Helsinki [139]. The Regional Committee on Research and Ethics (REK North) and the Norwegian Data Protection Authority (Datatilsynet) approved the Tromsø Study.

Participation in the study was voluntary, and all participants were informed about the study aims, types of data being collected, types of medical examinations, and how their results would be used in research. All the participants whose data is available in Tromsø7 gave written informed consent. The participants have a right to withdraw their consent at any time. Privacy and personal data protection are matters of ethical concern in medical research. In Tromsø7, all data presented in the research are anonymized using the following procedure: all participants in Tromsø7 were given personal identification numbers, “perskeys,” connected to their social security numbers. Then, another ID number was given to the “perskey” of each participant, so the “perskey” was not used in statistical analysis, providing an extra layer of data security.

For this project, we received separate approvals from REK North (reference number 68128) and the Norwegian Centre for Research Data (NSD) to use the data from the Tromsø Study database. We consider this project not bearing potential direct harm to the study participants. Nevertheless, all health data are sensitive, and unaccountable data handling can harm patients' privacy. According to the principles of medical research, patient privacy, and dignity should

be prioritized [139]. Therefore, we have familiarized ourselves with the Tromsø Study regulations on liable data handling. We stored the data in the Norwegian Service for Sensitive Data (TSD), owned by the University of Oslo, and operated and developed by the TSD service group at the University of Oslo, IT Department (USIT). The service is designed for storing and post-processing sensitive data in compliance with the Norwegian "Personal Data Act" and "Health Research Act" [140].

4 Results

4.1 Paper I

The specific aim of **paper I** was to calculate diagnostic efficacy parameters, such as sensitivity, specificity, accuracy, PPV, NPV, +LR, -LR, and AUC for the following indices and their combinations: MCW ≤3 mm, MCI-C2, C3, and MCI-C3 in females and males aged 40 years and older; MCW ≤3 mm, MCI-C2,C3, and MCI-C3 in females and males aged 65 years and older; MCW ≤3 mm, MCI-C2,C3, and MCI-C3 combined with FRAX scores > 15%.

Osteoporosis prevalence was 11.0% in females and 5.7% in males. The distributions of osteoporotic and non-osteoporotic cases by age, MCW, and MCI categories in females and males are presented in Appendix I and II, respectively. The distribution of T-scores by MCI categories is presented in Appendix III.

In females, severely eroded cortex (MCI-C3) and thin cortex (MCW ≤3 mm) showed the ability to distinguish between osteoporosis and non-osteoporosis as suggested by their positive and negative likelihood ratios and their confidence intervals, not including the value of 1. MCI-C3 showed the highest +LR of 3.3, meaning MCI-C3 was 3.3 times more likely to be present in females with osteoporosis compared to females without osteoporosis. In osteoporosis prediction, MCI-C3 and MCW ≤ 3mm showed a sensitivity of 34.0% and 68.1%, respectively. MCI-C3, i.e., severe cortical erosion, showed the highest specificity of 89.7% and was good at predicting non-osteoporotic cases. MCW ≤ 3mm showed a specificity of 69%. The AUC for MCW was 0.74 (95% CI, 0.67; 0.82).

When applying the same index tests to females aged 65 and older, MCI-C3 and MCW ≤ 3mm showed the ability to classify osteoporotic and non-osteoporotic cases. The sensitivities of both indices have increased at the expense of their specificities. The +LRs and -LR showed only a tiny change compared to the situation when MCI and MCW were applied to the population of females older than 40 years.

When the index tests were combined with FRAX scores > 15%, the predictive abilities of all indices were substantially improved. The combinations of FRAX with MCW or MCI at both thresholds (C3 and C2, C3) could differentiate between osteoporotic and non-osteoporotic females, as suggested by their likelihood ratio confidence intervals, not including 1. The sensitivities of all indices lowered while their specificities markedly increased. Severe cortical erosion and FRAX scores > 15% showed the highest predictive ability, demonstrating the +LR of 6.6, meaning that such a combination was 6.6 times more likely to be present in females with osteoporosis than healthy females.

Since the male sample was small and only 19 (5.7%) males had osteoporosis, the diagnostic efficacy estimates demonstrated wide confidence intervals and thus were inconclusive (Appendix IV).

4.2 Paper II

In this paper, the specific aim was to analyze general factors, such as T-score, age, and menopausal status (for females), and local factors, such as remaining teeth and periodontitis and their contribution to the morphology of mandibular cortex assessed by MCW, MCI, in females and males.

T-score, age, menopausal status, and the number of remaining teeth were significantly related to MCW in simple linear regression analyses in females. In multiple linear regression analysis with all factors added to the regression model, T-score, age, and number of teeth remained significant. T-score contributed most to the prediction of MCW in females, followed by age and remaining teeth with standardized β-coefficients of 0.286, -0.231, and 0.131, respectively. The model could predict MCW in females ($F=18.7$, $p<0.001$) and explained 24% of the variation in MCW.

In males, only the T-score was significantly related to MCW in simple linear regression analysis. However, in multiple regression analysis, none of the factors showed a significant association with MCW. The model could not significantly predict MCW in males ($F=2.3$, $p=0.059$).

When analyzing MCI in binary logistic regression analysis, T-score, age, menopausal status, and remaining teeth were significantly related to MCI in females. In multiple logistic regression, after adding all variables in the model, T-score, age, and remaining teeth remained significantly associated with MCI. Similarly to the results of linear regression analysis, the T-score contributed most to the thin and eroded cortex in females, followed by age and remaining teeth with the Wald values of 9.65, 6.17, and 5.83, respectively. The whole model was significant (Chi-square = 53.87, $p<0.001$) and could explain 16.3-23.0% of the variation in MCI.

In males, T-score, age, and remaining teeth were significantly related to MCI in binary linear regression analyses. After adding all variables in the model, only the T-score remained significantly related to MCI. The whole model could significantly predict MCI (Chi-square=12.01, $p=0.017$) and explained a 4.3-5.8% variation in MCI.

4.3 Paper III

This paper explored the feasibility of locating MF using deep learning of pre-trained object detectors and tailoring them to this specific task. The other aim was to explore the feasibility of combining the best-performing object detector with the algorithm measuring mandibular cortical width.

The second setup demonstrated better mAP for all four models than the setup I. EfficientDet-D0 demonstrated the best mAP at both IoU thresholds. The threshold of IoU = 0.5 was judged sufficient to consider the bounding box in the correct position, locating the mental foramen. With a 0.5 IoU threshold and setup II, the mAP of EfficientDet-D0 was 0.79, meaning that 79% of bounding boxes produced by the model were overlapping with the expert's annotations by more than 50%. The RetinaNet model showed the lowest average precision of 0.64 at IoU=0.5. When experts visually inspected the predictions of MF made by EfficientDet-D0 on handpicked «easy images,» they agreed on all cases (n=100). Notably, all IoU values between the algorithm predictions and the expert annotations, as well as confidence scores, were more than 0.5, confirming that the location of MF was an easy task. When experts inspected the predictions on 101 ”complex images,” they agreed with the algorithm on 67 images.

When assessing the reliability of the “ground truth,” the average IoU between the two experts was 0.678, with minimum and maximum values of 0.500 and 0.678, respectively.

Combining EficientDet-D0, the detector locating mental foramen, and the algorithm for MCW measurement was feasible. However, the algorithm should be further developed. The combined algorithm was tested on 100 images and produced output 93 times, 20 of which were unacceptable and measured other anatomical structures than the mandibular cortex.

5 Discussion

5.1 Methodological considerations

This thesis consists of three studies with a cross-sectional study design, i.e., all the data were collected simultaneously, and all study participants were recruited regardless of their exposure or disease status [141]. **Paper I** had the research question regarding diagnostic efficacy; **Paper II** was an exploratory study investigating the relationship between radiomorphometric indices general and local factors that may define cortical morphology, while **Paper III** aimed to estimate the technical efficacy of a new computer-aided tool. Cross-sectional study design is often used to study a disease's prevalence, analyze multiple outcomes, or study relationships between exposure and outcome variables. The main advantage of a cross-sectional study design is relatively inexpensive data collection. The disadvantage of this study design is temporal bias, i.e., the impossibility of defining the direction of the relationships between exposure and outcome and making any inferences about causality since all the data are collected simultaneously. Nevertheless, cross-sectional studies help to build theoretical models based on associations and test them further using other study designs [142]. Thus, the cross-sectional study design is suitable for all studies in this thesis, yet for answering the research question in **Paper I**, a prospective longitudinal study design would also fit and provide superior evidence.

Internal validity is the degree to which the results established in a study cannot be explained by other factors, and thus, these results are valid for the study population. External validity refers to the degree to which study results can be applied to a broader population [141, 143]. Internal validity is an essential prerequisite for external validity. Besides the temporal bias mentioned above, bias inherent to most epidemiological studies can threaten a cross-sectional

study's internal and external validity. These include random and non-random errors, also called bias, typically selection bias, information bias, and confounding [143].

5.1.1 Selection bias and external validity

Selection bias occurs when people have different chances to be included in a study depending on their exposure and outcome characteristics [141]. Selection bias threatens external validity in crossectional studies, and it is critical when the external validity of effect estimates is crucial, for example, in diagnostic efficacy studies, which are often cross-sectional. The study population in a diagnostic efficacy study should represent the population to which an index test should be applied [144, 145].

MCW and MCI in **Paper I** should preferably be generalizable to the patients of Norwegian dental clinics who can be susceptible to osteoporosis. Generally, people over 50 are considered a typical risk group [14]. Nevertheless, the disease develops earlier in females, specifically those experiencing menopause [146]. According to the World Health Organisation, menopause occurs between 45 and 55 years of age [147]. However, hip fracture incidence rates begin to increase linearly at the age of 40 in females [146]. Since there is a slight overlap in the age at which people become susceptible to osteoporosis, we considered people over 40 an eligible group for our study. Moreover, the FRAX tool allows calculations of fracture risk from 40 years of age [37].

In **Paper I**, participants were selected from a populational-based study. DEXA and DPRs in Tromsø7 were collected for a broad range of scientific purposes, including studies of disease prevalences for which representativeness is crucial. Participants who received both DEXA and DPRs were not deliberately selected for these examinations in Tromsø7, and thus, we can consider the study sample in **Paper I** to be random. However, a concern can be raised that

those chosen for **Paper I** may have a different prevalence of osteoporosis compared to all participants who received DEXA in Tromsø7. Figures 11 and 12 show the distributions of osteoporosis in females and males by age groups in our study participants compared to all Tromsø7 participants who were examined with DEXA. The total osteoporosis prevalence among all participants was 10.1% in females and 7.5% in males, while in our study, 11% and 5.7%, respectively. Thus, there is only a slight difference in osteoporosis prevalence between our study participants and all Tromsø7 participants. It holds for prevalence in all ages and by age groups. This difference is not critical, at least for females. A similar prevalence of osteoporosis at femoral necks was found in Tromsø5, carried out in 2001-2002, with a 79% response rate [11].

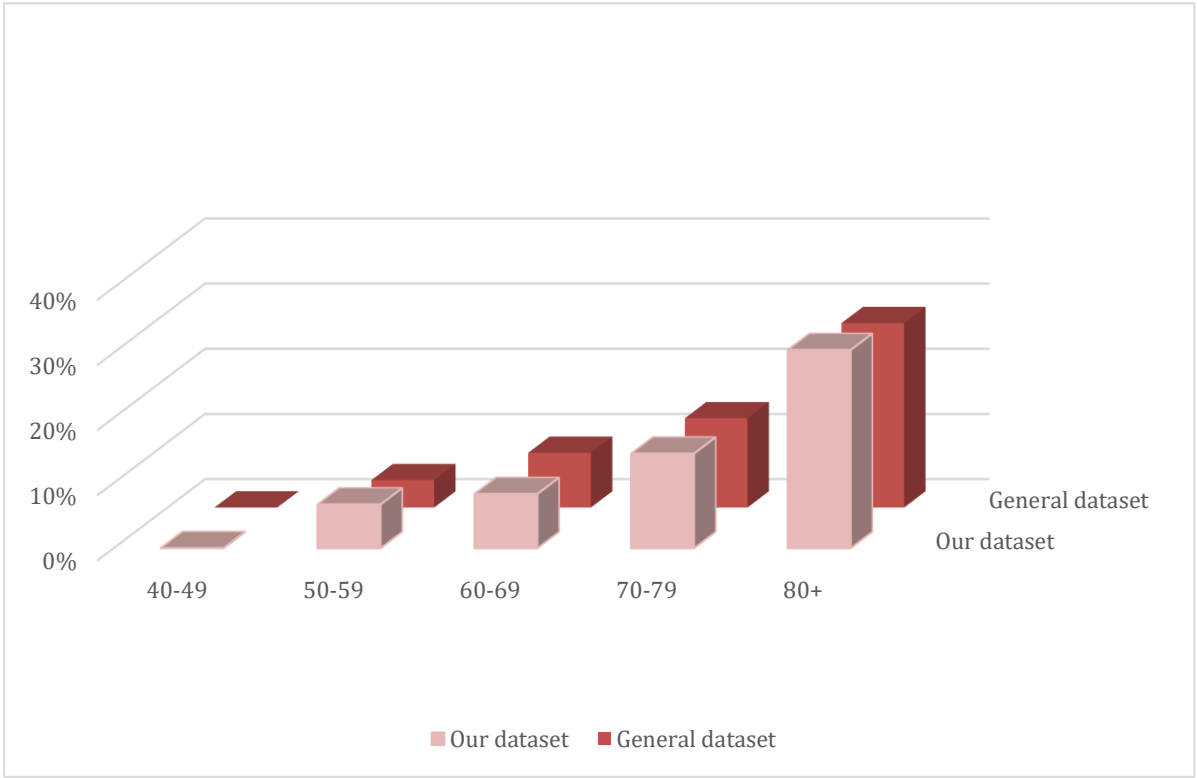


Figure 11 Distribution of osteoporosis in females in general Tromø7 dataset vs. our dataset

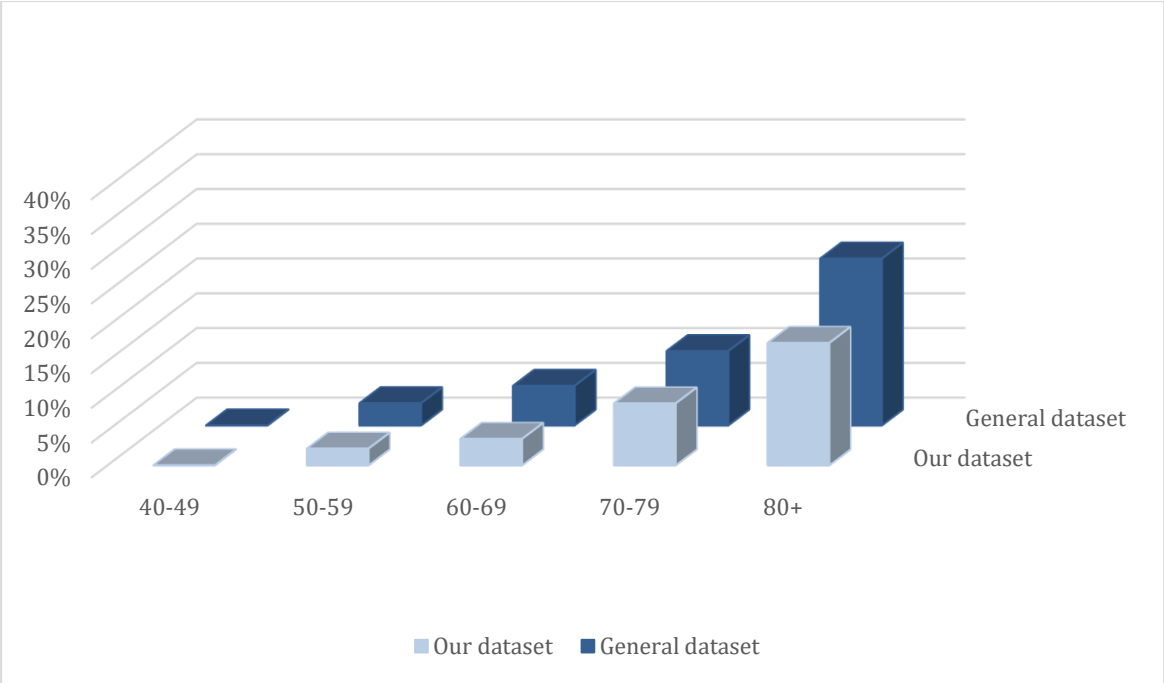


Figure 12 Distribution of osteoporosis in males in general Tromsø7 dataset vs. our dataset

Inappropriate exclusions, such as exclusions of "difficult-to-diagnose" cases or using disease cases vs. healthy controls verified from the study's beginning, can also introduce selection bias and compromise the external validity of a diagnostic efficacy study [145]. Such flaws in the sampling procedure produce a difference in test result distributions among participants with the disease compared to healthy ones, making two distributions further apart (Figure 13 A, B) and resulting in the inflation of sensitivity and specificity estimates [144, 145]. In **Paper I**, only underexposed or overexposed radiographs on which the mandibular cortex was not distinguishable were excluded from the index assessment. Since the T-scores of study participants were unknown before the MCI and MCW measurements were carried out, nobody was included or excluded based on the reference standard and osteoporosis status.

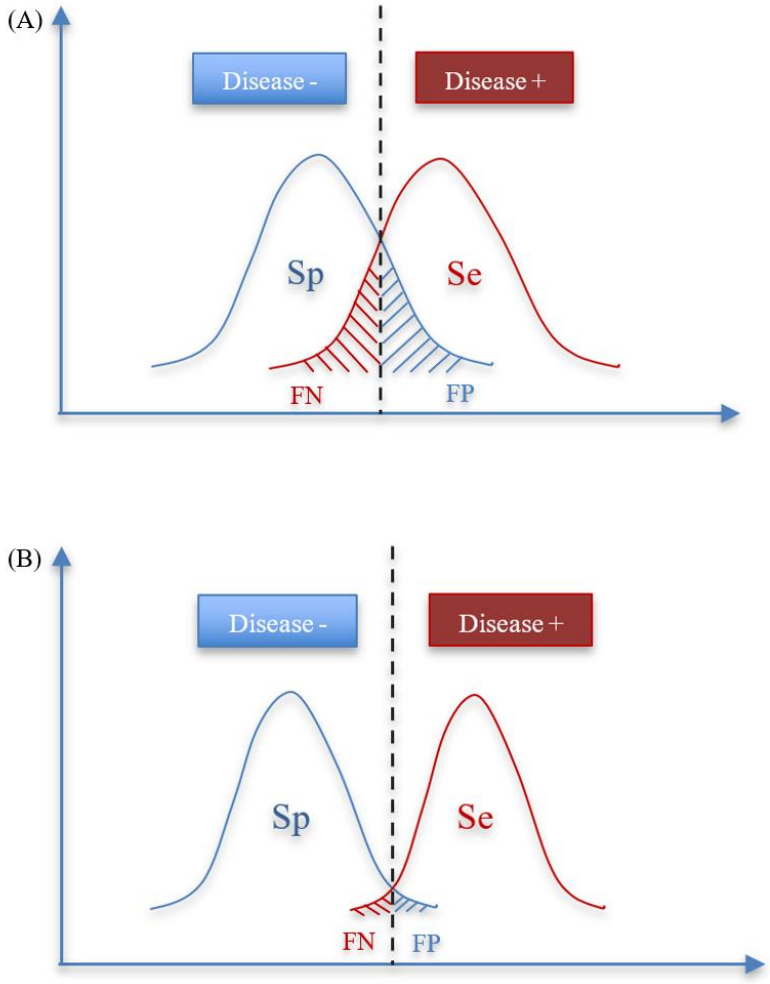


Figure 13 Distributions of test results among people with and without disease, sensitivities (Se), specificities (Sp), false positive (FP), and false negative (FN) test results. A) test result distributions among healthy and diseased people are closer to each other in a randomly selected population due to the presence of the whole severity spectrum, including borderline cases. B) In a non-randomly selected population, the test result distributions are further apart from each other due to inappropriate exclusions of borderline cases, the inclusion of known disease cases and healthy controls, or wrong settings with more severe disease cases

Besides sensitivity and specificity being sensitive to disease distribution in a particular population, some efficacy estimates such as PPV, NPV, and accuracy depend highly on disease prevalence [135]. Thus, merging males and females in one group for analysis is also a

source of selection bias. Differences in distributions of T-scores and osteoporosis prevalence in females and males were addressed by calculating diagnostic efficacy estimates separately.

Tromsø7 showed an attendance rate of 65%, which is high for a populational-based study. However, a concern that can be raised is that attendees may differ from non-attendees because they are healthier and can come for physical examination. However, this concern can be balanced with the fact that attendees of dental clinics are also more likely to be healthier than not attendees and be able to come to dental clinics physically.

Thus, selection bias was not of major concern in **Paper I**. The results of **Paper I** can be extrapolated at least on the population of Tromsø [148]. In **Paper II,** the same population was used. Had the sample been drawn from a more severely sick population, such as an orthopedic hospital, the inflated association estimates would have been observed, and the external validity would have been compromised. Although selection bias is less critical for **Paper II** since the estimates do not provide critical clinical information, the absence of selection bias strengthens the external validity of **Paper II**.

Paper III is a multidisciplinary paper involving machine learning methods with their own methodology and validation procedures. To my knowledge, universal quality guidelines for machine learning use in clinical and epidemiological research have not been launched yet. However, similar quality principles can also be applied to some extent to this type of study. For example, selection bias and external validity are common problems; results provided by a machine learning algorithm are mostly generalizable to the population with the same statistical distribution of data or image features as the population on which the algorithm has been trained. Shifts in data distribution may occur due to populational differences, differences in image acquisition protocols, and different manufacturers of medical imaging units [149]. T

To reach an optimal external validity of a machine learning algorithm, one needs to train it on data from a representative population and the images taken from different radiographic units using various protocols as it occurs in actual medical practice [150, 151]. Thus, the fact that DPRs for the training dataset were chosen from a populational-based study with a wide age range, anatomical variations, and image features strengthens the generalizability of the study results in **Paper III**. On the other hand, DPRs were taken from one panoramic machine, which weakens the result generalizability.

Furthermore, the algorithm locating mental foramen was prone to errors inherent to observers making the annotations from which the algorithm learns. Speculating how inconsistencies in automated location mental foramen affect inconsistencies in automated MCW measurements is difficult. However, this problem might be overcome by involving several well-trained observers for image annotation or using other methods not involving observers.

5.1.2 Information bias and internal validity

Information bias, or misclassification, is caused by flawed definitions or data collection of study variables, i.e., exposure and outcome. Misclassification can be systematic and non-systematic[141]. Systematic misclassification can compromise the internal validity of the study. Systematic misclassification in **papers I and II** could occur when measuring radiomorphometric indices. If MCI and MCW were measured knowing patients' T-scores, the raters would be more prone to over- or underestimate the thickness and erosion of the mandibular cortex depending on the participant’s osteoporosis status. Had such systematic misclassifications occurred in a diagnostic efficcay study, like **Paper I**, inflated efficacy estimates would have been obtained [152]. The same type of misclassification would result in exaggerated correlation and regression coefficients in a cross-sectional study exploring

associations between exposure and outcome like in **Paper II** [141]. Knowing that the raters in this project should be blind to T-scores, the radiomorphometric indices were first measured on DPRs and then sent to the Tromsø Study database administrators, who linked the newly recorded MCW and MCI with T-scores and the other Tromsø7 data and then sent it back for further analysis.

In diagnostic efficacy studies, reference standards can also be a source of information bias. First, knowledge of index test results by raters can influence reference standard measurements, leading to an exaggeration of the diagnostic efficacy estimates [145]. It was not the case in our **Paper I** because T-scores were measured in 2015-2016 during data collection in Tromsø7, i.e., before MCW and MCI data acquisition. Another essential assumption of a diagnostic efficacy study is that the reference standard correctly classifies a disease., i.e., has 100% sensitivity, and thus, all disagreements between index test and reference standard stem from incorrect index test results [145]. Since the T-score measured by DEXA at the femoral neck is an internationally recognized method for osteoporosis diagnosis [22], there is no concern regarding misclassifications caused by inappropriate reference standards in **Paper I.**

Another potential source of information bias in **Paper I** might be the FRAX score. While age, sex, and BMI data used to be reliable, frequency, amount of alcohol consumption, and smoking tend to be underreported in health surveys depending on sex, age, and social context [153, 154]. The data on the age at menopause are prone to recall bias., i.e., participants may hardly recall information from a distant past [141]. The data on rheumatoid arthritis, diabetes, or other diseases that can be secondary causes of osteoporosis were also self-reported. Self-reports on chronic diseases in populational-based studies are generally considered reliable [155]. The data on family history of hip fractures and information on previous fractures were

retrieved from Tromsø6, which causes significant concern regarding information bias when calculating the FRAX-score in **Paper I**. It is 7-8 years between the surveys of The Tromsø Study, and we expect some incident fractures to occur during this period, which may result in an underestimation of the FRAX score for some participants.

Non-systematic misclassification is a flawed exposure or outcome classification occurring randomly without a specific pattern. Non-systematic misclassification can decrease the precision of results and thus compromise the internal validity of a study. Study results, in that case, tend towards the null hypothesis, i.e., no relationships or associations [141 pp.187-193]. An example of a non-systematic misclassification is a measurement error. It is known from previous research that the DEXA procedure, performed by a well-calibrated staff, is reproducible and shows a 2% coefficient of variation between observers or in repeated observations. However, poor positioning can result in 5% errors at the femoral neck [24].

Random error in MCW and MCI occurs from the same sources: variation between observers or repeated observations and poor patient positioning. It raises a substantial concern in both **Paper I** and **Paper II**. We found suboptimal interobserver and intraobserver agreements [156]. When measuring MCW, the participants with the eroded cortex (C2 or C3) were the most complex cases because the upper border of the mandibular cortex consisted of trabecular layers and, therefore, was unclear.

Moreover, panoramic radiography is a complex procedure carried out by an operator. Positioning a patient in a panoramic machine is crucial for obtaining the correct projection of anatomical structures on a panoramic radiograph. Even though a patient is ideally positioned, the anatomical structures are projected with a magnified size. Magnification is inherent to panoramic radiography. Different panoramic machines have different magnification rates,

which are sometimes uneven in different image parts. A slight patient malposition was common in our dataset. A previous study reported that minor toward or backward shifts ±5 mm and tilts ±5° in the position of a patient result in less than a 2% coefficient of variation in vertical measurements on DPRs, which they considered clinically acceptable [157].

Nevertheless, only unreadable radiographs were excluded from our study, while some of the included radiographs revealed substantial malpositioning. It implies that the measurements of mandibular cortical width do not precisely reflect the actual thickness of the mandibular cortex, even when corrected for magnification rate. The measurement imprecisions result in random misclassifications of the thin cortex (thinner than 3 mm), noise in data, and underestimated efficacy [141]. Error related to malpositioning could have been avoided by using a reference object. Poor intro and interobserver agreement can be improved either by more extended training or by developing automatic algorithms measuring MCW. An attempt to develop such an algorithm has been made in **Paper III**.

5.1.3 Confounding, moderation, and internal validity

Confounding occurs when a variable is causally associated with the outcome and can be causally or non-causally associated with the exposure but cannot be variable on the causal pathway from exposure to outcome (Fig.14) [141 p.233].

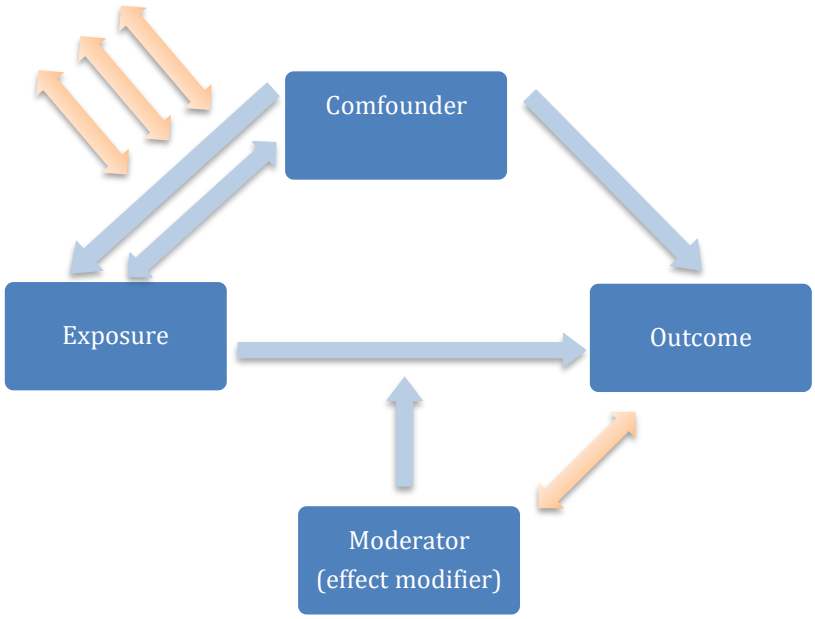


Figure 14 Mediation vs. moderation.

Paper II was an exploratory study. The primary objective was to find predictors associated with thin and eroded mandibular cortex and assess the strength of their associations. Generally, this type of study allows various possible predictors to be included in a multivariable regression model adjusted for general confounders such as age or sex, although the included predictors may have their confounders [158]. As in many epidemiological studies, age was a confounder in **Paper II** because it affected radiomorphometric indices and other variables such as T-score, menopausal status, remaining teeth, and periodontal disease. All analyses in **Paper II** were adjusted for age. However, the results of **Paper II** should be considered with caution: interpreting the regression coefficients as if the variables would be adjusted for each other is likely to result in the "table-2" fallacy because each variable may also have its own set of confounders that the multivariable model does not account for [159]. The latter will likely decrease the internal validity of the regression coefficients reported in **Paper II.**

A modifier (moderator) is a variable that changes the strength and direction of the relationships between exposure and outcome (Fig. 14) [143]. Based on previous studies [76, 78], we suspected that sex is a modifier, and it interacts with age. Therefore, we decided to separate analyses for males and females in **Paper II**. Had this interaction been not considered, it would have threatened the internal validity of the study results. Our study confirmed the interaction by demonstrating the difference in association strength between age and MCW and also found a slight difference in the association between MCW and T-score in males and females (Fig 15).

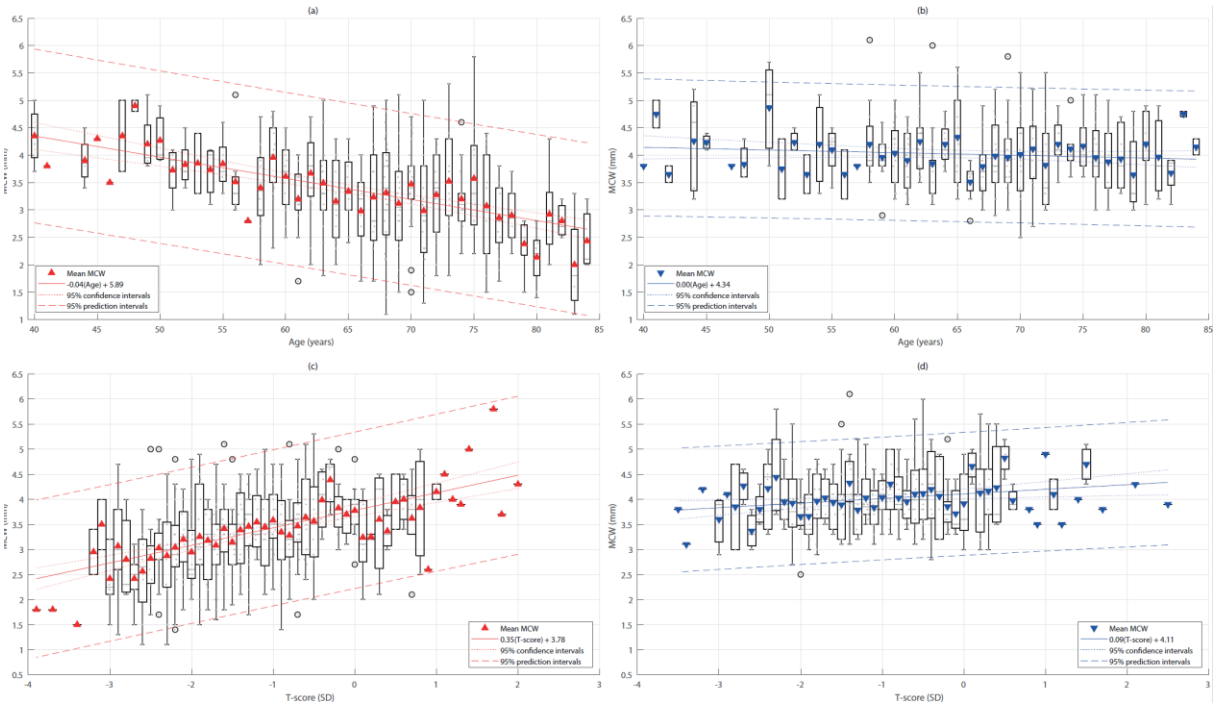


Figure 15 Box plots show mean MCW (filled triangles), its confidence intervals (boxes), and its distribution (vertical lines) in different ages in females (a) and males (b) and for different T-scores in females (c) and males (d). Solid lines are regression lines

5.1.4 Sample size considerations and precision

The best practice is to calculate sample size before initiating a study because carrying out a study with insufficient precision may be unethical, given that the participants receive

unnecessary medical examinations. However, only a post hoc sample size assessment can be done in our study since we used existing data.

The sample size calculation for diagnostic efficacy studies, like **Paper I**, is based on expected efficacy estimates (sensitivity and specificity) and desired two-sided confidence intervals [135]. A small sample size results in an inflated false discovery rate in diagnostic efficacy studies and imprecision of efficacy estimates with wide confidence intervals, rendering study results inconclusive and challenging for decision-making [160]. Generally, several factors would require a larger sample size: smaller estimates of sensitivity and specificity, lower disease prevalence, and narrower confidence intervals desired [161]. Since MCW and MCI are supposed to be screening tests, obtaining both high sensitivity and high specificity is not required. A sensitivity of approximately 70% combined with a specificity of at least 50% or vice versa is considered sufficient [160, 161]. We assumed that a 10% marginal error would be acceptable. Given the prevalence of osteoporosis and lower efficacy estimates in **Paper I,** it is evident that the precision of some efficacy estimates, especially sensitivity and likelihood ratios, is low, and marginal errors reach 15% from both sides in females. At the same time, the confidence intervals for specificity, accuracy, NPVs, and AUC in females are more precise, with a marginal error of 5%. All efficacy estimates showed low precision in males (Table 1 in appendices). At the same time, the World Health Organization guidelines for tests or diagnostic tools state that wide confidence intervals may not necessarily lead to downgrading evidence of a diagnostic efficacy study; if disease prevalence is low, then imprecise estimates, specifically sensitivity, might not be of particular concern [162p 228].

Linear and logistic regression analyses were used in **Paper II**, which requires different sample size calculations. Various online tools allow sample size calculations for linear and logistic regression analysis based on ultimate characteristics such as power, significance, type

of statistical test, and expected effect size. Since we could not use an expected effect size based on previous research, we used a well-established rule of thumb from epidemiologic literature for both regression analyses. Green et al. reviewed several rules of thumb and concluded that regression analysis requires at least 200 subjects. For linear regression analysis, they concluded that a minimum of subjects per independent variable should be 15-25 [163]. In **Paper II**, 427 females and 335 males were included in linear regression analysis with MCW as an outcome, while the total number of predictors was 5. Peduzzi et al. suggested the minimum number of events per variable; they stated that the number of subjects in the smallest group with a particular outcome should be a minimum of 10 per each independent variable [164, 165]. In **Paper II**, the smallest group for MCI outcome was the participants with C3 cortex. There were 55 females and one male; therefore, we merged this group with C2, which included 243 females and 136 males. There were five predictors in logistic regression analysis. Thus, there is no concern regarding sample size and precision of estimates in **Paper II**.

There are no clear guidelines regarding the sample size needed for machine learning model application in medical imaging analyses. Therefore, we did not carry out sample size calculations in **Paper III**. A recent systematic review has raised this issue, stating that sample size requirements are needed to develop machine-learning algorithms for medical purposes [166]. Another problem affecting sample size in medical imaging studies using machine learning is that medical images cannot be publicly available as other images used for training datasets due to patient privacy and security policies [123]. Therefore, obtaining sufficient samples of medical images for training, testing, and validation is challenging. Generally, it has been proposed that the more data in the training dataset, the better the accuracy of the model can be achieved [167]. Classic statistical methods for sample size calculations have

drawbacks because they do not consider specific characteristics of AI modeling [168]. However, some researchers use the Widrow-Hoff learning rule, which suggests obtaining data from 10 patients for every imaging feature the model uses [168]. Balki et al. suggested different analytical approaches based on algorithm characteristics or curve-fitting approaches, empirically exploring model performance at different sample sizes [166]. Generally, several factors affect the required sample size for the machine-learning algorithm. First is the number of features used in the model and their variability: the more features and the larger the variability between them, the more images are needed. The second factor is the number of observers annotating images in supervised machine learning algorithms: the fewer observers and the broader variations between their annotations, the more images are needed [169]. Similarly to plain epidemiological studies, disease prevalence and clinically relevant effect sizes to distinguish between healthy and diseased patients also play roles in sample size calculations in medical imaging studies using AI. The quality of medical images varies significantly in real life, requiring more images for model training. Thousands of cases are typically required for accurate model training [169].

5.2 Discussion of the main results

The usefulness of a potential screening test is a complex multi-level issue. Besides the diagnostic qualities and reproducibility of the radiomorphometric indices explored in **Paper I**, it was also essential to understand better the link between radiomorphometric indices, skeletal BMD, and other factors potentially affecting this link, which has not been much studied before and which was done in **Paper II**. Exploring the feasibility of automated measurements of mandibular cortical morphology in **Paper III** is a further step towards standardizing the test conduct, improving its reproducibility, and saving time for dental practitioners.

5.2.1 Factors affecting morphology of mandibular cortex

Paper II explored several factors affecting mandibular cortical morphology in a large study population. The results of **Paper II** showed that the T-score was the factor contributing most to the morphology of the mandibular cortex in females. Our findings were consistent with previous studies showing that females with osteoporosis had thinner and more eroded cortices than healthy females [89, 90, 170, 171]. Age was another factor significantly associated with radiomorphometric in females, which was in line with the results from previous studies [76, 78, 81]. **Paper II** also found that the more teeth remained, the thicker and less eroded cortex was observed in females. The relation between remaining teeth and mandibular cortex morphology appeared more controversial: some previous studies found a link between remaining teeth and MCW and MCI [80, 82], while others did not find it for MCW [78, 81]. Menopausal status and periodontal disease were not associated with the morphology of the mandibular cortex in females in our study, unlike previous studies that found an association [87, 88].

Interestingly, none of the factors were associated with mandibular cortical morphology in males except the T-score related to MCI with an OR of 0.73 (95% CI 0.56; 0.96). A 0.1 SD higher T-score reduced the odds of having mildly (C2) or severely eroded cortex (C3) by 27% or 0.73 times. Relationships between radiomorphometric indices and skeletal T-score reflecting osteoporosis status have been poorly studied in males. Two studies reported significant differences in MCW and MCI in osteoporotic vs. non-osteoporotic males, but the adjustments were lacking [104, 106]. In our study, age in multivariable regression was not significantly associated with MCW or MCI, while another study found a slight decrease in MCW with age in males, yet less prominent than in females [76]. No association between

remaining teeth and radiomorphometric indices in males was found in this study, similar to the results reported by another study [80]. One may speculate that the differences in factors affecting the mandibular cortex in males and females might be related to the differences in bone physiology, specifically bone remodeling [44]. It is known that sexual dimorphism in bone morphology develops in puberty during growth, and sex hormones affect the remodeling of bone throughout life [44, 45]. In older ages, the resorption rates of the endosteal cortex are similar in both sexes, while periosteal bone apposition is faster in males than in females [44]. The studies exploring the geometry of the femoral neck and other parts of the femur also found differences between males and females in the thickness of the cortex with older age [172, 173].

5.2.2 Efficacy of radiomorphometric indices and their utility for osteoporosis screening

The clinical usefulness of radiomorphometric indices can be defined by their diagnostic efficacy. The diagnostic efficacy results for MCW and MCI in males in **Paper I** were inconclusive due to the low prevalence of osteoporosis and wide confidence intervals. However, **Paper II's** findings suggest that the utility of using MCW and MCI in males is likely to be low since the T-score was poorly associated with the morphology of the mandibular cortex in males. However, further studies are needed to confirm it. At the same time, the results of **Paper I** showed that MCW $\leq$ 3 mm and MCI (C3) could distinguish between osteoporosis and non-osteoporosis in females. Our study's AUC for female MCW was 0.74 (95%, CI 0.67; 0.82), demonstrating moderate efficacy [174]. The US Preventive Task Force, in their recommendation statement on osteoporosis screening, reported similar AUCs for peripheral DEXA (AUC from 0.67 to 0.80), FRAX (AUC from 0.58 to 0.82), for

ORAI, OSIRIS, OST, and SCORE (AUC from 0.65 to 0.70) for detection of osteoporosis in females older than 40 years [35]. Sensitivities and specificities for MCW and MCI alone in the population older than 40 years were similar to those reported in the previous systematic review [90].

Applying MCW and MCI to the population aged 65 years and older improved their sensitivities at the expense of specificities but did not improve the positive and negative LRs. The trade-off between sensitivity and specificity in older females might be attributed to the higher osteoporosis prevalence in older age and the differences in the T-score spectrum. Although there is no mathematical relationship between sensitivity, specificity, and disease prevalence, Leeflang et al. showed, using a series of examples from systematic reviews, that specificity and specificity covary with prevalence through the distribution of other clinical variables [175, 176]. Combining MCW or MCI with FRAX > 15% improved their specificities, yet sensitivities have somewhat decreased. It is a known phenomenon that the sensitivity of two tests combined with the "AND" rule (the diagnosis is positive when both tests are positive) will be lower than the sensitivities of each separate test, while the specificity will be higher [135]. In addition, the combination of radiomorphometric indices with FRAX showed superior diagnostic efficacy compared with radiomorphometric indices alone, as suggested by the increase in +LRs and relatively stable -LR. The highest +LR of 6.6 was observed for the combination of MCI-C3 and FRAX, indicating a moderate ability of the two tests combined to predict osteoporosis [177]. Only one study used a similar combination of radiomorphometric index with the FRAX score, combining sparse trabeculation on intraoral radiographs and FRAX > 15% to predict osteoporotic fractures. Similar to our results, they found a superior predictive ability for the combination of two tests than for sparse trabeculation alone [178]. In their study, the predictive ability of sparse trabeculation

combined with FRAX was comparable to that of FRAX combined with BMD to predict osteoporotic fractures [178].

The diagnostic efficacy estimates do not usually have a straightforward interpretation facilitating a decision on whether the test is useful. None of the efficacy estimates alone are interpretable; one should understand them in combination [162]. A higher sensitivity or a higher specificity might be of value in different circumstances. There are also several aspects to consider. The first aspect is whether a new test should replace the existing diagnostic standard or be used as an add-on or a triage test before another diagnostic procedure. In the first case, one wants high sensitivity and specificity, while the latter does not require it. The second aspect to consider is whether further diagnostics or effective treatment is available [162]. The tests under evaluation in this project, MCW and MCI, are triage tests and should be taken to inform the decision about the need for further DEXA examination, which means that both high sensitivity and specificity are not required. DEXA is a well-established diagnostic test [22], and effective therapies preventing osteoporotic fractures also exist [1]. Another aspect to consider is the consequences of false-positive and false-negative results for patients, such as unnecessary diagnostics and treatment or the absence of treatment when needed [162]. The harms of osteoporosis screening with DEXA have not been well studied. However, the United States Preventive Services Task Force reviewed existing evidence and concluded that there were no potential psychological harms related to unnecessary diagnostics since the examination is not invasive and minor potential harms related to false-positive results and unnecessary treatment [35]. Thus, the arguments above support using MCW and MCI for osteoporosis prescreening.

Recommendations on diagnostic test use are rarely based solely on diagnostic efficacy estimates [162]. A diagnostic test produces some clinical information that does not

necessarily improve patients' health. Thus, the further aspects of a new diagnostic test, such as its impact on health outcomes and cost-effectiveness, should be considered [162, 179]. In the case of osteoporosis screening, we should further study in randomized control trials (RCTs) whether a new test and a subsequent preventive treatment would change osteoporotic fracture incidence in patients. Cost-effectiveness is another relevant aspect of test utility that is highly dependent on the country's context. The knowledge of the abovementioned aspects is lacking for MCW and MCI, which limits the judgment about the usefulness of MCW and MCI for osteoporosis screening.

5.2.3 Developing a fully automated algorithm for MCW measurement

Even though **Paper II** found the strongest association of indices with T-score among other factors, and **Paper I** found the ability of radiomorphometric indices to predict osteoporosis in females, the suboptimal agreement of the indices between observers and repeated observations would disfavor the use of MCI and MCW for osteoporosis screening. Our findings regarding inter and intraobserver agreement were similar to the results of a previous study [108]. Sutthiprapaporn et al. attempted to solve this problem through extensive observer training [180], while the others focused on developing an automated algorithm measuring MCW or MCI on DPRs [112, 113, 119].

Devlin et al. developed a method using active shape models (ASM) or so-called «snakes» [113, 181]. They reported higher reproducibility and diagnostic efficacy of this method than the manual one. However, they found that manual initiation of ASM gave better results than fully automated [113]. The drawback of this method is that ASM was sensitive to image quality and irregularities in the mandibular cortex, such as overlapping anatomical structures, having trouble finding cut-offs between different intensity values, and, thus, ASM was not

robust. Moreover, fully automated ASM could not identify MCW in the particular anatomical projection, for example, mental foramen [113]. Allen et al. also developed an algorithm measuring MCW, but they manually defined the location of mental foramen [112]. Both studies concluded that developing a fully automatic method would be beneficial because osteoporosis is not a primary pathology in dental clinics, and minimal time should be spent on such incidental yet significant findings [112, 113].

In our **Paper III,** we developed a fully automatic algorithm measuring MCW based on fine-tuned pre-trained neural network locating mental foramen and combined another automated algorithm based on grayscale thresholding measuring mandibular cortical width. We considered the mean average precision of mental foramen location of 0.79 to be satisfying because the experts agreed with all suggestions made by the algorithm on 100 easy images and most suggestions on 101 complex ones, though IoUs with experts' annotations were sometimes $< 50\%$. MCW measurements produced in this study were not compared with the patient's T-score; therefore, the efficacy of the MCW method is unknown. Aliaga et al. attempted to solve the problem of automatic detection of anatomical landmarks, including mental foramen, used for radiomorphometric indices measurements [182]. Unlike ours, their algorithm was based on thresholding and the assumption that mental foramen is a dark area with specific gray values, specific size, and other features. They reported that the method failed in 5% of 310 cases [182]. Similar to Aliaga [182], our method was aimed to imitate the original one proposed by Ledgerton et al. [81]. However, alternative automatic methods of analyzing mandibular bone quality can automatically assess osteoporosis status based on panoramic radiographs. For example, Lee et al. used a transfer learning strategy on a convolutional neural network [183]. This method does not require finding anatomical landmarks; the algorithm extracts the necessary features independently. For example, in the

study by Lee et al., the features extracted by the algorithm were the low border of the mandibular cortex and spongious bone on its periphery, which support the idea that the morphology of the mandibular cortex may reveal the osteoporotic status of a patient [183].

5.3 Future perspectives:

Given that the Tromsø Study collects DPRs, DEXA, and other clinical information for many scientific purposes, conducting a more extensive study would be worthwhile. Tromsø Study can be linked to the Norwegian hip fracture registry [148]. Thus, it would be feasible to carry out a longitudinal study using panoramic indices alone or combined with FRAX to predict hip fracture.

Moreover, the Trøndelag Health Study (the HUNT study) collects similar populational-based data on DEXA and DPRs for various scientific purposes [184]. Tromsø study and HUNT study can be combined [185], which allows a large generalizable sample that can be used to obtain more precise efficacy estimates. An extensive study sample based on the Tromsø study and HUNT would also allow the development of a more robust artificial intelligence algorithm classifying patients as osteoporotic or healthy. For example, an algorithm based on unsupervised machine learning would not require annotation made by dental experts and thus avoids common human errors. It has also been recognized that the wider the variety of images taken from different image modalities that the algorithm is trained on, the better its generalizability [151].

Even though our findings suggest a poor association of T-score with the morphology of the mandibular cortex and, thus, potentially low utility of radiomorphometric indices in males, further studies are needed to confirm or contradict it.

5.4 Conclusions:

In this project, we found that radiomorphometric indices, MCW, and MCI are related to skeletal BMD to a great extent, and they can differentiate between osteoporosis and non-osteoporosis in females, demonstrating moderate efficacy similar to the currently used risk-factor-based tools. Combining MCW and MCI with FRAX substantially improves their diagnostic efficacy, specifically their ability to detect healthy females.

Skeletal BMD appears weakly associated with mandibular cortical morphology in males, but further studies on the utility of radiomorphometric indices are needed.

The reliability of manual MCW and MCI assessments is suboptimal, but we found it feasible to create a machine learning algorithm that automatically measures MCW. The algorithm accurately locates the mental foramen, the initial and most challenging step in measuring MCW, and can be combined with another algorithm measuring the thickness of the mandibular cortex. Two algorithms together allow fully automated measurement of MCW, but the method should be further improved and validated against participants' osteoporosis status.

Works cited

1. Screening to Prevent Osteoporotic Fractures: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. . [Internet]. 2018 [cited 9/19/2022]. Report No.: 0098-7484 Contract No.: 24. Available from: https://doi.org/10.1001/jama.2018.6537

2. Hauge IHR, Widmark A, Bruzell EM. Bruk av røntgendiagnostikk blant norske tannlegar : prosjektretta tilsyn etter ny forskrift om strålevern og bruk av stråling. Østerås: Statens Strålevern; 2009.

3. Radiation use in Norway. Useful use and good radiation protection for society, humans and the environment. [Internet]. 2014 [cited. Available from: https://dsa.no/publikasjoner/stralevernrapport-2-2014-stralebruk-i-norge/StralevernRapport_02-2014_Str%C3%A5lebruk%20i%20Norge.pdf

4. Helseforhold, levekårsundersøkelsen: Bruk av tannlege, fysioterapeut og andre helsetjenester (prosent), etter kjønn, alder, statistikkvariabel, år og type helseproblem Atatistisk sentralbyrå; 2012 [cited 2023 Apr 24]. Available from: https://www.ssb.no/statbank/table/07261/tableViewLayout1/.

5. Gülcan F, Nasir E, Klock KS, Eckbäck G, Åstrøm AN. Sosiale forskjeller i tannhelse og bruk av tannhelsetjenester fra 65-til 70 års alder-et livsløpsperspektiv. Den norske tannlegeforenings Tidende. [Internet]. 2014 [cited 2021 Dec 15]; 192:[99 p.]. Available from: https://www.tannlegetidende.no/asset/2014/P14-03-192-9.pdf.

6. Vernekar N, Batchelor P, Heilmann A. Adult self-reported attendance for dental check-ups over a 16-year period in the UK. British Dental Journal. [Internet]. 2019 [cited 2021 Dec 8]; 226(11):[883-8 pp.]. Available from: https://doi.org/10.1038/s41415-019-0366-8.

7. Organisation WH. World Health Organization scientific group on the assessment of osteoporosis at primary health care level. Summary Meeting Report Brussels, Belgium, 5-7 May 2004. [Internet]. 2007 [cited. Available from: https://www.dur-a-avaler.com/wp-content/uploads/2015/04/WHO-2004-Osteoporosis.pdf

8. Søgaard AJ, Gustad TK, Bjertness E, Tell GS, Schei B, Emaus N, et al. Urban-rural differences in distal forearm fractures: Cohort Norway. Osteoporosis International. [Internet]. 2007 [cited 2021 Dec 8]; 18(8):[1063-72 pp.]. Available from: https://doi.org/10.1007/s00198-007-0353-9.

9. 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). Archives of Osteoporosis. 2013;8(1-2):1-115 http://dx.doi.org/10.1007/s11657-013-0136-1

10. Sarafrazi N, Wambogo EA, Shepherd JA. Osteoporosis or low bone mass in older adults: United States, 2017–2018. [Internet]. 2021. Available from: https://dx.doi.org/10.15620/cdc:103477.

11. Emaus N, Omsland TK, Ahmed LA, Grimnes G, Sneve M, Berntsen GK. Bone mineral density at the hip in Norwegian women and men—prevalence of osteoporosis depends on chosen references: the Tromsø Study. European Journal of Epidemiology. [Internet]. 2009 [cited 2022 June 9]; 24(6):[321-8 pp.]. Available from: https://doi.org/10.1007/s10654-009-9333-z.

12. Eduard MH. Facts about osteoporosis and osteoporosis fractures Oslo: Norwegian Institute of Public Health; 2004 [updated 20.10.2016; cited. Available from: https://www.fhi.no/fp/folkesykdommer/beinskjorhet/beinskjorhet-og-brudd---fakta-om-os/.

13. Stasistics: population Statistics Norway; 2009 [cited. Available from: https://www.ssb.no/statbank/table/07459/tableViewLayout1/.

14. Epidemiology of osteoporosis and fragility fractures 9, rue Juste-Olivier CH-1260 Nyon - Switzerland: International Osteoporotic Society; [cited. Available from: https://www.osteoporosis.foundation/facts-statistics/epidemiology-of-osteoporosis-and-fragility-fractures.

15. Willers C, Norton N, Harvey NC, Jacobson T, Johansson H, Lorentzon M, et al. Osteoporosis in Europe: a compendium of country-specific reports. Archives of Osteoporosis. 2022;17(1):23 http://dx.doi.org/10.1007/s11657-021-00969-8

16. Downey C, Kelly M, Quinlan JF. Changing trends in the mortality rate at 1-year post hip fracture - a systematic review. World Journal of Orthopedics. 2019;10(3):166-75 http://dx.doi.org/10.5312/wjo.v10.i3.166

17. Kanis JA, Norton N, Harvey NC, Jacobson T, Johansson H, Lorentzon M, et al. SCOPE 2021: a new scorecard for osteoporosis in Europe. Archives of Osteoporosis. [Internet]. 2021 [cited 2023 Feb 21]; 16(1):[82 p.]. Available from: https://doi.org/10.1007/s11657-020-00871-9.

18. Odén A, McCloskey EV, Kanis JA, Harvey NC, Johansson H. Burden of high fracture probability worldwide: secular increases 2010-2040. Osteoporosis International. 2015;26(9):2243-8 http://dx.doi.org/10.1007/s00198-015-3154-6

19. Tønnessen M. Aldring i Norge. Sosiologi i dag Aldring, omsorg of familier. [Internet]. 2015; 45(3). Available from: https://ojs.novus.no/index.php/SID/article/view/1169.

20. Hagen G, Magnussen J, Tell G, Omsland T. Estimating the future burden of hip fractures in Norway. A NOREPOS study. Bone. 2020;131:115156 http://dx.doi.org/10.1016/j.bone.2019.115156

21. Hendrickson NR, Pickhardt PJ, Del Rio AM, Rosas HG, Anderson PA. Bone mineral density T-scores derived from CT attenuation numbers (Hounsfield units): clinical utility and correlation with dual-energy X-ray absorptiometry. The Lowa Orthopaedic Journal. [Internet]. 2018 PMC6047377]; 38:[25 p.]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6047377/.

22. Kanis JA, Cooper C, Rizzoli R, Reginster J-Y. European guidance for the diagnosis and management of osteoporosis in postmenopausal women. Osteoporosis International. [Internet]. 2019 [cited 2021 Sep 1]; 30(1):[3-44 pp.]. Available from: https://dx.doi.org/10.1007%2Fs00198-018-4704-5.

23. Lewiecki EM, Gordon CM, Baim S, Leonard MB, Bishop NJ, Bianchi M-L, et al. International Society for Clinical Densitometry 2007 adult and pediatric official positions. Bone. 2008;43(6):1115-21 http://dx.doi.org/10.1016/j.bone.2008.08.106

24. Adams JE. Advances in bone imaging for osteoporosis. Nature Reviews Endocrinology. 2013;9(1):28-42

25. Crandall CJ, Newberry SJ, Diamant A, Lim YW, Gellad WF, Booth MJ, et al. Comparative effectiveness of pharmacologic treatments to prevent fractures: an updated systematic review. Annals of Internal Medicine 2014;161(10):711-23 http://dx.doi.org/10.7326/m14-0317

26. Crandall CJ, Newberry SJ, Diamant A, Lim YW, Gellad WF, Suttorp MJ, et al. Treatment To Prevent Fractures in Men and Women With Low Bone Density or Osteoporosis: Update of a 2007 Report. [Internet]. 2012. Available from: https://pubmed.ncbi.nlm.nih.gov/22553885/.

27. Marshall D, Johnell O, Wedel H. Meta-analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures. The BMJ. 1996;312(7041):1254-9 http://dx.doi.org/10.1136/bmj.312.7041.1254

28. Ström O, Borgström F, Kanis JA, Compston J, Cooper C, McCloskey EV, et al. Osteoporosis: burden, health care provision and opportunities in the EU: a report prepared in collaboration with the International Osteoporosis Foundation (IOF) and the European Federation of Pharmaceutical Industry Associations (EFPIA). Archives of Osteoporosis. 2011;6:59-155 http://dx.doi.org/10.1007/s11657-011-0060-1

29. Overview of places in Norway where bone mineral density examination is available Norwegian osteoporosis association; 2020 [cited. Available from: https://nof-norge.no/wp-content/uploads/2020/02/dexa_maalesteder_april_2020.pdf.

30. Manca A, Willan AR. 'Lost in translation': accounting for between-country differences in the analysis of multinational cost-effectiveness data. Pharmacoeconomics. 2006;24(11):1101-19 http://dx.doi.org/10.2165/00019053-200624110-00007

31. Swedish Council on Health Technology A. SBU Systematic Review Summaries. Osteoporosis – Prevention, Diagnosis and Treatment: A Systematic Review. Stockholm: Swedish Council on Health Technology Assessment (SBU)

32. Joakimsen RM, Fønnebø V. Screening for osteoporosis in Norway? Tidsskrift for Den norske legeforening. [Internet]. 2001 [cited 2022 Sep 11]; 121(5):[615-8 pp.]. Available from: https://tidsskriftet.no/2001/02/kronikk/bor-det-innfores-screening-osteoporose-i-norge.

33. Turner DA, Khioe RFS, Shepstone L, Lenaghan E, Cooper C, Gittoes N, et al. The cost‐effectiveness of screening in the community to reduce osteoporotic fractures in older women in the UK: economic evaluation of the SCOOP study. Journal of Bone and Mineral Research. 2018;33(5):845-51 http://dx.doi.org/10.1002/jbmr.3381

34. Yu EW. Screening for osteoporosis in postmenopausal women and men. UpToDate. [cited 2021 Dec 12]. Available from: https://www.uptodate.com/contents/2046.

35. U.S. Preventive Services Task Force. Screening for Osteoporosis to Prevent Fractures. US Preventive Services Task Force Recommendation Statement.U.S. Preventive Services Task Force. 2018; Mar 1):[cited 2023, Mar 1] Mar 1.Available from: https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/osteoporosis-screening

36. Kanis JA, Oden A, Johansson H, Borgström F, Ström O, McCloskey E. FRAX® and its applications to clinical practice. Bone. [Internet]. 2009 [cited 2021 Oct 22]; 44(5):[734-43 pp.]. Available from: https://doi.org/10.1016/j.bone.2009.01.373.

37. FRAX® Fracture Risk Assessment Tool Centre for Metabolic Bone Diseases, University of Sheffield, UK; 2008 [cited 2021 Dec 22]. Available from: https://www.sheffield.ac.uk/FRAX/index.aspx.

38. Eriksen EF. Osteoporosis. Norwegian electronic handbook for medical doctors. Norwegian Health Institute; 2021. [cited 2021 Dec 22]. Available from: https://legehandboka.no/handboken/kliniske-kapitler/endokrinologi/tilstander-og-sykdommer/beinvev-og-vitamin-d/osteoporose.

39. Hoff M, Skurtveit S, Meyer HE, Langhammer A, Søgaard AJ, Syversen U, et al. Anti-osteoporosis drug use: too little, too much, or just right? The HUNT study, Norway. Osteoporosis International. 2018;29(8):1875-85 http://dx.doi.org/10.1007/s00198-018-4560-3

40. The Norwegian Society for rheumatology: osteoporosis: investigation, treatment and follow-up The Norwegian Society for rheumatology:; 2015 [cited. Available from: http://legeforeningen.no/PageFiles/229641/Osteoporose%20%202015.pd.

41. Baig M, Bacha D. Histology, Bone. StatPearls [Internet]. [Internet]. 2022. Available from: https://www.ncbi.nlm.nih.gov/books/NBK541132/.

42. Clarke B. Normal Bone Anatomy and Physiology. Clinical Journal of the American Society of Nephrology. 2008;3(Supplement_3):S131-S9 http://dx.doi.org/10.2215/cjn.04151206

43. Jonasson G, Skoglund I, Rythén M. The rise and fall of the alveolar process: Dependency of teeth and metabolic aspects. Archives of Oral Biology. 2018;96:195-200 http://dx.doi.org/10.1016/j.archoralbio.2018.09.016

44. Seeman E. Sexual dimorphism in skeletal size, density, and strength. The Journal of Clinical Endocrinology & Metabolism. 2001;86(10):4576-84 http://dx.doi.org/10.1210/jcem.86.10.7960

45. Callewaert F, Sinnesael M, Gielen E, Boonen S, Vanderschueren D. Skeletal sexual dimorphism: relative contribution of sex steroids, growth hormone-insulin-like growth factor-I (GH-IGF-I) and mechanical loading. Journal of Endocrinology. 2010;207(2):127-34 http://dx.doi.org/10.1677/joe-10-0209

46. Lanyon LE. Functional strain in bone tissue as an objective, and controlling stimulus for adaptive bone remodelling. Journal of Biomechanics. 1987;20(11-12):1083-93 http://dx.doi.org/10.1016/0021-9290(87)90026-1

47. Miller CJ, Trichilo S, Pickering E, Martelli S, Delisser P, Meakin LB, et al. Cortical thickness adaptive response to mechanical loading depends on periosteal position and varies linearly with loading magnitude. Frontiers in Bioengineering and Biotechnology. 2021;9:671606 http://dx.doi.org/10.3389/fbioe.2021.671606

48. Lang T, LeBlanc A, Evans H, Lu Y, Genant H, Yu A. Cortical and trabecular bone mineral loss from the spine and hip in long-duration spaceflight. Journal of Bone and Mineral Research. 2004;19(6):1006-12 http://dx.doi.org/10.1359/jbmr.040307

49. Groen J, Duyvensz F, Halsted J. Diffuse alveolar atrophy of the jaw (non-inflammatory form of paradental disease) and pre-senile osteoporosis. Gerontologia Clinica. 1960;2(2):68-86 http://dx.doi.org/10.1159/000244610

50. Hildebolt CF. Osteoporosis and oral bone loss. Dentomaxillofacial Radiology. 1997;26(1):3-15 http://dx.doi.org/10.1038/sj.dmfr.4600226

51. Krook L, Whalen J, Lesser G, Lutwak L. Human periodontal disease and osteoporosis. The Cornell Veterinarian. [Internet]. 1972; 62(3):[371-91 pp.]. Available from: https://pubmed.ncbi.nlm.nih.gov/5039605/.

52. Henrikson P-å, Lutwak L, Krook L, Skogerboe R, Kallfelz F, Bélanger LF, et al. Fluoride and nutritional osteoporosis: physicochemical data on bones from an experimental study in dogs. The Journal of Nutrition. 1970;100(6):631-42 http://dx.doi.org/10.1093/jn/100.6.631

53. Atkinson P, Woodhead C. Changes in human mandibular structure with age. Archives of Oral Biology. 1968;13(12):1453-IN16 http://dx.doi.org/10.1016/0003-9969(68)90027-7

54. Manson J, Lucas R. A microradiographic study of age changes in the human mandible. Archives of Oral Biology. 1962;7(6):761-IN14 http://dx.doi.org/10.1016/0003-9969(62)90125-5

55. Von Wowern N, Stoltze K. Sex and age differences in bone morphology of mandibles. Scandinavian Journal of Dental Research. 1978;86(6):478-85 http://dx.doi.org/10.1111/j.1600-0722.1978.tb00654.x

56. Von Wowern N. Bone mass of mandibles. In vitro and in vivo analyses. Danish Medical Bulletin. [Internet]. 1986; 33(1):[23-44 pp.]. Available from: https://pubmed.ncbi.nlm.nih.gov/3948537/.

57. Bays RA, Weinstein RS. Systemic bone disease in patients with mandibular atrophy. Journal of Oral and Maxillofacial Surgery. 1982;40(5):270-2 http://dx.doi.org/10.1016/0278-2391(82)90216-6

58. Rosenquist JB, Baylink DJ, Berger JS. Alveolar atrophy and decreased skeletal mass of the radius. International Journal of Oral Surgery. 1978;7(5):479-81 http://dx.doi.org/10.1016/s0300-9785(78)80040-4

59. Goldberg A, Gergans G, Mattson D, Rudman D. Radiographic alveolar process/mandibular height ratio as a predictor of osteoporosis. Gerodontics. 1988;4(5):229-31

60. Bras J, van Ooij CP, Abraham-Inpijn L, Wilmink JM, Kusen GJ. Radiographic interpretation of the mandibular angular cortex: a diagnostic tool in metabolic bone loss. Part II. Renal osteodystrophy. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology. 1982;53(6):647-50 http://dx.doi.org/10.1016/0030-4220(82)90356-5

61. Bras J, van Ooij CP, Abraham-Inpijn L, Kusen GJ, Wilmink JM. Radiographic interpretation of the mandibular angular cortex: A diagnostic tool in metabolic bone loss. Part I. Normal state. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology. 1982;53(5):541-5 http://dx.doi.org/10.1016/0030-4220(82)90473-x

62. Henrikson P-Å, Wallenius, Kjell. The mandible and osteoporosis (1) A qualitative comparison between the mandible and the radius. Journal of Oral Rehabilitation. [Internet]. 1974; 1(1):[67-74 pp.]. Available from: https://pubmed.ncbi.nlm.nih.gov/4528045/.

63. Von Wowern N, Storm TL, Olgaard K. Bone mineral content by photon absorptiometry of the mandible compared with that of the forearm and the lumbar spine. Calcified Tissue International. 1988;42:157-61 http://dx.doi.org/10.1007/BF02556328

64. Blake GM, Fogelman I. The role of DXA bone density scans in the diagnosis and treatment of osteoporosis. Postgraduate Medical Journal. 2007;83(982):509-17 http://dx.doi.org/10.1136/pgmj.2007.057505

65. Hildebolt C, Rupich R, Vannier M, Zerbolio Jr D, Shrout M, Cohen S, et al. Inter‐relationships between bone mineral content measures: Dual energy radiography (DER) and bitewing radiographs (BWX). Journal of Clinical Periodontology. 1993;20(10):739-45 http://dx.doi.org/10.1111/j.1600-051x.1993.tb00700.x

66. Ledgerton D, Horner K, Devlin H, Worthington H. Panoramic mandibular index as a radiomorphometric tool: An assessment of precision. Dentomaxillofacial Radiology. [Internet]. 1997 [cited 2021 Nov 25]; 26(2):[95-100 pp.]. Available from: https://doi.org/10.1038/sj.dmfr.4600215.

67. Benson BW, Prihoda TJ, Glass BJ. Variations in adult cortical bone mass as measured by a panoramic mandibular index. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology. 1991;71(3):349-56 http://dx.doi.org/10.1016/0030-4220(91)90314-3

68. Klemetti E, Kolmakov S, Kröger H. Pantomography in assessment of the osteoporosis risk group. Scandinavian Journal of Dental Research. [Internet]. 1994; 102(1):[68-72 pp.]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/8153584.

69. Lindh C, Petersson A, Rohlin M. Assessment of the trabecular pattern before endosseous implant treatment: diagnostic outcome of periapical radiography in the mandible. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology. 1996;82(3):335-43 http://dx.doi.org/10.1016/s1079-2104(96)80363-5

70. Yeung AWK, Mozos I. The innovative and sustainable use of dental panoramic radiographs for the detection of osteoporosis. International Journal of Environmental Research and Public Health. [Internet]. 2020 [cited 2021 Dec 16]; 17(7). Available from: https://www.ncbi.nlm.nih.gov/pubmed/32260243.

71. Taguchi A, Ohtsuka M, Tsuda M, Nakamoto T, Kodama I, Inagaki K, et al. Risk of vertebral osteoporosis in post-menopausal women with alterations of the mandible. Dentomaxillofacial Radiology. 2007;36(3):143-8 http://dx.doi.org/10.1259/dmfr/50171930

72. Devlin H, Horner K. Mandibular radiomorphometric indices in the diagnosis of reduce skeletal bone mineral density. Osteoporosis International. [Internet]. 2002 [cited 2021 Nov 12]; 13(5):[373-8 pp.]. Available from: https://doi.org/10.1007/s001980200042.

73. Taguchi A, Suei Y, Ohtsuka M, Otani K, Tanimoto K, Ohtaki M. Usefulness of panoramic radiography in the diagnosis of postmenopausal osteoporosis in women. Width

and morphology of inferior cortex of the mandible. Dentomaxillofacial Radiology. [Internet]. 1996 [cited 2021 Nov 10]; 25(5):[263-7 pp.]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/9161180.

74. Marandi S, Bagherpour A, Imanimoghaddam M, Hatef MR, Haghighi AR. Panoramic-Based Mandibular Indices and Bone Mineral Density of Femoral Neck and Lumbar Vertebrae in Women. Frontiers in dentistry. [Internet]. 2010; 7(2). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3184742/.

75. Okabe S, Morimoto Y, Ansai T, Yoshioka I, Tanaka T, Taguchi A, et al. Assessment of the relationship between the mandibular cortex on panoramic radiographs and the risk of bone fracture and vascular disease in 80-year-olds. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology. 2008;106(3):433-42 http://dx.doi.org/10.1016/j.tripleo.2007.09.013

76. Roberts M, Yuan J, Graham J, Jacobs R, Devlin H. Changes in mandibular cortical width measurements with age in men and women. Osteoporosis International. [Internet]. 2011 [cited 2022 Jan 9]; 22(6):[1915-25 pp.]. Available from: https://doi.org/10.1007/s00198-010-1410-3.

77. Kavitha MS, Park SY, Heo MS, Chien SI. Distributional variations in the quantitative cortical and trabecular bone radiographic measurements of mandible, between male and female populations of korea, and its utilization. PloS One. [Internet]. 2016 [cited 2021 Dec 17]; 11(12):[e0167992 p.]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28002443.

78. Dutra V, Yang J, Devlin H, Susin C. Radiomorphometric indices and their relation to gender, age, and dental status. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology. 2005;99(4):479-84 http://dx.doi.org/10.1016/j.tripleo.2004.09.013

79. Al-Dam A, Haydar Y, Rashad A. Is there a change in the mandibular cortical shape index in panoramic radiographs over time? A longitudinal retrospective study. JDental Research Journal. [Internet]. 2020 [cited The authors of this manuscript declare that they have no conflicts of interest, real or perceived, financial or non-financial in this article. PMC8045519]; 17(6):[465-71 pp.]. Available from: https://pubmed.ncbi.nlm.nih.gov/33889353/.

80. Taguchi A, Tanimoto K, Suei Y, Wada T. Tooth loss and mandibular osteopenia. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology. 1995;79(1):127-32 http://dx.doi.org/10.1016/s1079-2104(05)80088-5

81. Ledgerton D, Horner K, Devlin H, Worthington H. Radiomorphometric indices of the mandible in a British female population. Dentomaxillofacial Radiology. [Internet]. 1999 [cited 2021 Nov 25]; 28(3):[173-81 pp.]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/10740473.

82. Gulsahi A, Yuzugullu B, Imirzalıoglu P, Genç Y. Assessment of panoramic radiomorphometric indices in Turkish patients of different age groups, gender and dental status. Dentomaxillofacial Radiology. 2008;37(5):288-92 http://dx.doi.org/10.1259/dmfr/19491030

83. Hienz SA, Paliwal S, Ivanovski S. Mechanisms of bone resorption in periodontitis. Journal of immunology research. 2015;2015 http://dx.doi.org/10.1155/2015/615486

84. Demmer RT, Trinquart L, Zuk A, Fu BC, Blomkvist J, Michalowicz BS, et al. The influence of anti-infective periodontal treatment on C-reactive protein: a systematic review and meta-analysis of randomized controlled trials. PloS One. 2013;8(10):e77441 http://dx.doi.org/10.1371/journal.pone.0077441

85. Souza PP, Lerner UH. The role of cytokines in inflammatory bone loss. Immunological Investigations. 2013;42(7):555-622 http://dx.doi.org/10.3109/08820139.2013.822766

86. Martínez-Maestre M, González-Cejudo C, Machuca G, Torrejón R, Castelo-Branco C. Periodontitis and osteoporosis: a systematic review. Climacteric. 2010;13(6):523-9 http://dx.doi.org/10.3109/13697137.2010.500749

87. Kolte RA, Kolte AP, Potey AM. Risk assessment of osteoporosis in pre- and postmenopausal periodontally healthy and chronic periodontitis women with digital panoramic radiographs. Journal of Indian Society of Periodontology. 2017;21(6):461-5 http://dx.doi.org/10.4103/jisp.jisp_238_17

88. Mudda JA, Bajaj M, Patil VA. A Radiographic comparison of mandibular bone quality in pre- and post-menopausal women in Indian population. Journal of Indian Society of Periodontology. 2010;14(2):121-5 http://dx.doi.org/10.4103/0972-124x.70833

89. Calciolari E, Donos N, Park JC, Petrie A, Mardas N. Panoramic measures for oral bone mass in detecting osteoporosis: a systematic review and meta-analysis. Journal of Dental Research. [Internet]. 2015 [cited 2021 Dec 20]; 94(3 Suppl):[17S-27S pp.]. Available from: https://doi.org/10.1177/0022034514554949.

90. Kinalski MA, Boscato N, Damian MF. The

of panoramic radiography as a screening of bone mineral density in women: a systematic review. Dentomaxillofacial Radiology. [Internet]. 2020 [cited 2021 Dec 14]; 49(2):[20190149 p.]. Available from: https://doi.org/10.1259/dmfr.20190149.

91. Taguchi A, Sanada M, Krall E, Nakamoto T, Ohtsuka M, Suei Y, et al. Relationship between dental panoramic radiographic findings and biochemical markers of bone turnover. Journal of Bone and Mineral Research. 2003;18(9):1689-94 http://dx.doi.org/10.1359/jbmr.2003.18.9.1689

92. Ardakani FE, Owlia MB, Hesami S, Hosseini P. Digital panoramic radiography as a useful tool for detection of bone loss: a comparative study. Acta Medica Iranica. [Internet]. 2013:[94-100 pp.]. Available from: https://pubmed.ncbi.nlm.nih.gov/23585315/.

93. Devlin H, Karayianni K, Mitsea A, Jacobs R, Lindh C, van der Stelt P, et al. Diagnosing osteoporosis by using dental panoramic radiographs: the OSTEODENT project. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology. 2007;104(6):821-8 http://dx.doi.org/10.1016/j.tripleo.2006.12.027

94. Karayianni K, Horner K, Mitsea A, Berkas L, Mastoris M, Jacobs R, et al. Accuracy in osteoporosis diagnosis of a combination of mandibular cortical width measurement on dental panoramic radiographs and a clinical risk index (OSIRIS): the OSTEODENT project. Bone. [Internet]. 2007 [cited 2022 Jan 2]; 40(1):[223-9 pp.]. Available from: https://doi.org/10.1016/j.bone.2006.07.025.

95. Nicopoulou-Karayianni K, Tzoutzoukos P, Mitsea A, Karayiannis A, Tsiklakis K, Jacobs R, et al. Tooth loss and osteoporosis: the OSTEODENT Study. Journal of Clinical Periodontology. 2009;36(3):190-7 http://dx.doi.org/10.1111/j.1600-051X.2008.01365.x

96. Taguchi A, Tsuda M, Ohtsuka M, Kodama I, Sanada M, Nakamoto T, et al. Use of dental panoramic radiographs in identifying younger postmenopausal women with osteoporosis. Osteoporosis International. 2006;17:387-94 http://dx.doi.org/10.1007/s00198-005-2029-7

97. Taguchi A, Tanimoto K, Suei Y, Ohama K, Wada T. Relationship between the mandibular and lumbar vertebral bone mineral density at different postmenopausal stages. Dentomaxillofacial Radiology. 1996;25(3):130-5 http://dx.doi.org/10.1259/dmfr.25.3.9084261

98. Taguchi A, Tanaka R, Kakimoto N, Morimoto Y, Arai Y, Hayashi T, et al. Clinical guidelines for the application of panoramic radiographs in screening for osteoporosis. Oral Radiology. 2021 http://dx.doi.org/10.1007/s11282-021-00518-6

99. Cosman F, de Beur SJ, LeBoff M, Lewiecki E, Tanner B, Randall S, et al. Clinician's guide to prevention and treatment of osteoporosis. Osteoporosis International. 2014;25:2359-81 http://dx.doi.org/10.1007/s00198-014-2794-2

100. Crandall CJ, Larson J, Gourlay ML, Donaldson MG, LaCroix A, Cauley JA, et al. Osteoporosis Screening in Postmenopausal Women 50 to 64 Years Old: Comparison of US Preventive Services Task Force Strategy and Two Traditional Strategies in the Women's Health Initiative. Journal of Bone and Mineral Research. [Internet]. 2014 [cited 2022 Sep 10]; 29(7):[1661-6 pp.]. Available from: https://doi-org.mime.uit.no/10.1002/jbmr.2174.

101. Horner K, Allen P, Graham J, Jacobs R, Boonen S, Pavitt S, et al. The relationship between the OSTEODENT index and hip fracture risk assessment using FRAX. Oral Surgery,

Oral Medicine, Oral Pathology, Oral Radiology. 2010;110(2):243-9 http://dx.doi.org/10.1016/j.tripleo.2010.03.035

102. Kalinowski P, Różyło-Kalinowska I, Piskórz M, Bojakowska-Komsta U. Correlations between periodontal disease, mandibular inferior cortex index and the osteoporotic fracture probability assessed by means of the fracture risk assessment body mass index tool. BMC Medical Imaging. 2019;19:1-8 http://dx.doi.org/10.1186/s12880-019-0337-1

103. Jonasson GB, Sundh V, Hakeberg M, Ahlqwist M, Lissner L, Hange D. Evaluation of clinical and radiographic indices as predictors of osteoporotic fractures: a 10-year longitudinal study. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology. 2018;125(5):487-94 http://dx.doi.org/10.1016/j.oooo.2017.11.009

104. Dagistan S, Bilge O. Comparison of antegonial index, mental index, panoramic mandibular index and mandibular cortical index values in the panoramic radiographs of normal males and male patients with osteoporosis. Dentomaxillofacial Radiology. 2010;39(5):290-4 http://dx.doi.org/10.1259/dmfr/46589325

105. Sindeaux R, de Souza Figueiredo PT, de Melo NS, Guimarães ATB, Lazarte L, Pereira FB, et al. Fractal dimension and mandibular cortical width in normal and osteoporotic men and women. Maturitas. 2014;77(2):142-8 http://dx.doi.org/10.1016/j.maturitas.2013.10.011

106. Leite AF, de Souza Figueiredo PT, Ramos Barra F, Santos de Melo N, de Paula AP. Relationships between mandibular cortical indexes, bone mineral density, and osteoporotic fractures in Brazilian men over 60 years old. 2011 [cited 2022 Jan 11]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21903429.

107. Taguchi A, Asano A, Ohtsuka M, Nakamoto T, Suei Y, Tsuda M, et al. Observer performance in diagnosing osteoporosis by dental panoramic radiographs: results from the osteoporosis screening project in dentistry (OSPD). Bone. 2008;43(1):209-13 http://dx.doi.org/10.1016/j.bone.2008.03.014

108. Devlin CV, Horner K, Devlin H. Variability in measurement of radiomorphometric indices by general dental practitioners. Dentomaxillofacial Radiology. 2001;30(2):120-5 http://dx.doi.org/10.1038/sj/dmfr/4600594

109. Horner K, Karayianni K, Mitsea A, Berkas L, Mastoris M, Jacobs R, et al. The mandibular cortex on radiographs as a tool for osteoporosis risk assessment: the OSTEODENT Project. Journal of Clinical Densitometry. 2007;10(2):138-46 http://dx.doi.org/10.1016/j.jocd.2007.02.004

110. Arifin AZ, Asano A, Taguchi A, Nakamoto T, Ohtsuka M, Tsuda M, et al. Computer-aided system for measuring the mandibular cortical width on dental panoramic radiographs in identifying postmenopausal women with low bone mineral density. Osteoporosis International. 2006;17:753-9 http://dx.doi.org/10.1007/s00198-005-0045-2

111. Arifin AZ, Asano A, Taguchi A, Nakamoto T, Ohtsuka M, Tsuda M, et al. Use of fuzzy neural network in diagnosing postmenopausal women with osteoporosis based on dental panoramic radiographs. Journal of Advanced Computational Intelligence and Intelligent Informatics. 2007;11(8) http://dx.doi.org/10.20965/jaciii.2007.p1049

112. Allen PD, Graham J, Farnell DJ, Marjanovic EJ, Adams J, Jacobs R, et al., editors. Detecting osteoporosis from dental radiographs using active shape models. 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 2007: IEEE.

113. Devlin H, Allen PD, Graham J, Jacobs R, Karayianni K, Lindh C, et al. Automated osteoporosis risk assessment by dentists: a new pathway to diagnosis. Bone. 2007;40(4):835-42 http://dx.doi.org/10.1016/j.bone.2006.10.024

114. Roberts M, Graham J, Devlin H, editors. Improving the detection of osteoporosis from dental radiographs using active appearance models. 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 2010: IEEE.

115. Devlin H, Allen P, Graham J, Jacobs R, Nicopoulou-Karayianni K, Lindh C, et al. The role of the dental surgeon in detecting osteoporosis: the OSTEODENT study. British Dental Journal. 2008;204(10):E16-E http://dx.doi.org/10.1038/sj.bdj.2008.317

116. Nakamoto T, Taguchi A, Ohtsuka M, Suei Y, Fujita M, Tsuda M, et al. A computer-aided diagnosis system to screen for osteoporosis using dental panoramic radiographs. Dentomaxillofacial Radiology. 2008;37(5):274-81 http://dx.doi.org/10.1259/dmfr/68621207

117. Sukegawa S, Fujimura A, Taguchi A, Yamamoto N, Kitamura A, Goto R, et al. Identification of osteoporosis using ensemble deep learning model with panoramic radiographs and clinical covariates. Scientific Reports. 2022;12(1):1-10 http://dx.doi.org/10.1038/s41598-022-10150-x

118. Kavitha MS, An S-Y, An C-H, Huh K-H, Yi W-J, Heo M-S, et al. Texture analysis of mandibular cortical bone on digital dental panoramic radiographs for the diagnosis of osteoporosis in Korean women. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology. 2015;119(3):346-56 http://dx.doi.org/10.1016/j.oooo.2014.11.009

119. Nakamoto T, Taguchi A, Kakimoto N. Osteoporosis screening support system from panoramic radiographs using deep learning by convolutional neural network. Dentomaxillofacial Radiology. 2022;51(6):20220135 http://dx.doi.org/10.1259/dmfr.20220135

120. Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njolstad I. Cohort profile: The Tromso Study. International Journal of Epidemiology 2012;41(4):961-7 http://dx.doi.org/10.1093/ije/dyr049

121. Hopstock LA, Grimsgaard S, Johansen H, Kanstad K, Wilsgaard T, Eggen AE. The seventh survey of the Tromsø Study (Tromsø7) 2015–2016: study design, data collection,

attendance, and prevalence of risk factors and disease in a multipurpose population-based health survey. Scandinavian Journal of Public Health. [Internet]. 2022 [cited 2022 Oct 17]; 50(7):[919-29 pp.]. Available from: https://doi.org/10.1177/14034948221092294.

122. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nature Methods. 2012;9(7):671-5 http://dx.doi.org/10.1038/nmeth.2089

123. Greenspan H, Van Ginneken B, Summers RM. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Transactions on Medical Imaging. 2016;35(5):1153-9 http://dx.doi.org/10.1109/TMI.2016.2553401

124. Dutta A, Zisserman A, editors. The VIA annotation software for images, audio and video. Proceedings of the 27th ACM international conference on multimedia; 2019.

125. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al., editors. Microsoft coco: Common objects in context. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13; 2014: Springer.

126. He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.

127. Newell A, Yang K, Deng J, editors. Stacked hourglass networks for human pose estimation. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14; 2016: Springer.

128. Tan M, Le Q, editors. Efficientnet: Rethinking model scaling for convolutional neural networks. International conference on machine learning; 2019: PMLR.

129. Lin T-Y, Goyal P, Girshick R, He K, Dollár P, editors. Focal loss for dense object detection. Proceedings of the IEEE international conference on computer vision; 2017.

130. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems arXiv.org e-Print archive; 2016 [cited. Available from: https://doi.org/10.48550/arXiv.1603.04467.

131. Edvardsen IP, Teterina A, Johansen T, Myhre JN, Godtliebsen F, Bolstad NL. Automatic detection of the mental foramen for estimating mandibular cortical width in dental panoramic radiographs: the seventh survey of the Tromsø Study (Tromsø7) in 2015–2016. Journal of International Medical Research. 2022;50(11) http://dx.doi.org/10.1177/03000605221135147

132. Holde GE, Oscarson N, Trovik TA, Tillberg A, Jönsson B. Periodontitis prevalence and severity in adults: a cross‐sectional study in Norwegian circumpolar communities. Journal of Periodontology. 2017;88(10):1012-22

133. Tonetti MS, Greenwell H, Kornman KS. Staging and grading of periodontitis: Framework and proposal of a new classification and case definition. Journal of Periodontology. 2018;89:S159-S72 http://dx.doi.org/10.1002/JPER.18-0006

134. Trombelli L, Farina R, Silva CO, Tatakis DN. Plaque‐induced gingivitis: Case definition and diagnostic considerations. Journal of Clinical Periodontology. 2018;45:S44-S67 http://dx.doi.org/10.1111/jcpe.12939

135. Zhou X-H, McClish DK, Obuchowski NA. Statistical Methods in Diagnostic Medicine. Hoboken: John Wiley & Sons; 2011.

136. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics. [Internet]. 1977 [cited 2021 Dec 22]; 33(1):[159-74 pp.]. Available from: https://pubmed.ncbi.nlm.nih.gov/843571/.

137. Padilla R, Netto SL, Da Silva EA, editors. A survey on performance metrics for object-detection algorithms. 2020 international conference on systems, signals and image processing (IWSSIP); 2020: IEEE.

138. Rezatofighi H, Tsoi N, Gwak J, Sadeghian A, Reid I, Savarese S, editors. Generalized intersection over union: A metric and a loss for bounding box regression. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2019.

139. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects.World Medical Association. JAMA. 2013; 310(20):2191-4. [cited 2021 Oct 13] 20.Available from: https://doi.org/10.1001/jama.2013.281053

140. About TSD: Citation of TSD University of Oslo; 2016 [updated Nov 18, 2021; cited. Available from: https://www.uio.no/english/services/it/research/sensitive-data/about/cite-tsd.html.

141. Szklo M, Nieto FJ. Epidemiology: Beyond the Basics. 4th ed. Sudbury: Jones & Bartlett Learning, LLC; 2018.

142. Mann C. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. Emergency Medicine Journal. 2003;20(1):54-60 http://dx.doi.org/10.1136/emj.20.1.54

143. Porta M. A Dictionary of Epidemiology. Oxford: Oxford: Oxford University Press, Incorporated; 2014.

144. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, Van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. Canadian Medical Association Journal. 2006;174(4):469-76 http://dx.doi.org/10.1503/cmaj.050090

145. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of Internal Medicine 2011;155(8):529-36 http://dx.doi.org/10.7326/0003-4819-155-8-201110180-00009

146. Cooper C, Melton LJ. Epidemiology of osteoporosis. Trends in Endocrinology and Metabolism. 1992;3(6):224-9 http://dx.doi.org/10.1016/1043-2760(92)90032-V

147. Menopause. Key facts World Health Organisation; 2022 [cited. Available from: https://www.who.int/news-room/fact-sheets/detail/menopause#:~:text=Most%20women%20experience%20menopause%20between,in%20circulating%20blood%20oestrogen%20levels.

148. Eggen AE, Mathiesen EB, Wilsgaard T, Jacobsen BK, Njølstad I. The sixth survey of the Tromsø Study (Tromsø 6) in 2007–08: Collaborative research in the interface between clinical medicine and epidemiology: Study objectives, design, data collection procedures, and attendance in a multipurpose population-based health survey. Scandinavian Journal of Public Health. 2013;41(1):65-80 http://dx.doi.org/10.1177/1403494812469851

149. Yu AC, Eng J. One algorithm may not fit all: how selection bias affects machine learning performance. Radiographics. 2020;40(7):1932-7 http://dx.doi.org/10.1148/rg.2020200040

150. Subbaswamy A, Saria S. From development to deployment: dataset shift, causality, and shift-stable models in health AI. Biostatistics. 2020;21(2):345-52 http://dx.doi.org/10.1093/biostatistics/kxz041

151. Artificial intelligence in healthcare. [Internet]. 2022 [cited Contract No.: PE 729.512. Available from: https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729512/EPRS_STU(2022)729512_EN.pdf

152. Rohlin M, Horner K, Lindh C, Wenzel A. Through the quality kaleidoscope: reflections on research in dentomaxillofacial imaging. Dentomaxillofacial Radiology. 2020;49(6):20190484

153. Connor Gorber S, Schofield-Hurwitz S, Hardt J, Levasseur G, Tremblay M. The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and cotinine-assessed smoking status. Nicotine & Tobacco Research. 2009;11(1):12-24 http://dx.doi.org/10.1093/ntr/ntn010

154. Livingston M, Callinan S. Underreporting in alcohol surveys: whose drinking is underestimated? Journal of studies on alcohol and drugs. [Internet]. 2015; 76(1):[158-64 pp.]. Available from: https://pubmed.ncbi.nlm.nih.gov/25486405/.

155. Najafi F, Moradinazar M, Hamzeh B, Rezaeian S. The reliability of self-reporting chronic diseases: how reliable is the result of population-based cohort studies. Journal of Preventive Medicine and Hygiene. 2019;60(4):E349-e53 http://dx.doi.org/10.15167/2421-4248/jpmh2019.60.4.1118

156. Teterina A, Niratisairak S, Morseth B, Bolstad N. Diagnostic efficacy of radiomorphometric indices for predicting osteoporosis in a Norwegian population in the Tromsø Study: Tromsø7. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology. 2022 http://dx.doi.org/10.1016/j.oooo.2022.10.039

157. Xie Q, Soikkonen K, Wolf J, Mattila K, Gong M, Ainamo A. Effect of head positioning in panoramic radiography on vertical measurements: an in vitro study. Dentomaxillofacial Radiology. 1996;25(2):61-6 http://dx.doi.org/10.1259/dmfr.25.2.9446974

158. Oxford University Press US. Is it a risk factor, a predictor, or even both? The multiple faces of multivariable regression analysis.Oxford University Press US. 2023; 188(1):lvac012. [cited 1.Available from:

159. Westreich D, Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. American Journal of Epidemiology. 2013;177(4):292-8 http://dx.doi.org/10.1093/aje/kws412

160. Van Stralen KJ, Stel VS, Reitsma JB, Dekker FW, Zoccali C, Jager KJ. Diagnostic methods I: sensitivity, specificity, and other measures of accuracy. Kidney International. 2009;75(12):1257-63 http://dx.doi.org/10.1038/ki.2009.92

161. Bujang MA. An Elaboration on Sample Size Planning for Performing a One-Sample Sensitivity and Specificity Analysis by Basing on Calculations on a Specified 95% Confidence Interval Width. Diagnostics. 2023;13(8):1390 http://dx.doi.org/10.3390/diagnostics13081390

162. Developing guideline recommendations for tests or diagnostic tools. WHO handbook for guideline development. World Health Organization; 2014 [cited 2022 Jan 13]. Available from: https://apps.who.int/iris/bitstream/handle/10665/145714/9789241548960-chap17-eng.pdf.

163. Green SB. How many subjects does it take to do a regression analysis. Multivariate behavioral research. 1991;26(3):499-510 http://dx.doi.org/10.1207/s15327906mbr2603_7

164. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. Journal of Clinical Epidemiology. 1996;49(12):1373-9 http://dx.doi.org/10.1016/S0895-4356(96)00236-3

165. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. Journal of Clinical Epidemiology. 1995;48(12):1503-10 http://dx.doi.org/10.1016/0895-4356(95)00048-8

166. Balki I, Amirabadi A, Levman J, Martel AL, Emersic Z, Meden B, et al. Sample-size determination methodologies for machine learning in medical imaging research: a systematic review. Canadian Association of Radiologists Journal. 2019;70(4):344-53 http://dx.doi.org/10.1016/j.carj.2019.06.002

167. Cho J, Lee K, Shin E, Choy G, Do S. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? arXiv.org e-Print archive; 2015 [cited. Available from: https://doi.org/10.48550/arXiv.1511.06348.

168. Castiglioni I, Rundo L, Codari M, Di Leo G, Salvatore C, Interlenghi M, et al. AI applications to medical images: From machine learning to deep learning. Physica Medica. 2021;83:9-24 http://dx.doi.org/10.1016/j.ejmp.2021.02.006

169. Harvey H, Glocker B. A Standardised Approach for Preparing Imaging Data for Machine Learning Tasks in Radiology. In: Ranschaert ER, Morozov S, Algra PR, editors. Artificial Intelligence in Medical Imaging: Opportunities, Applications and Risks. Cham: Springer International Publishing; 2019. p. 61-72.

170. Leite AF, Figueiredo PT, Guia CM, Melo NS, de Paula AP. Correlations between seven panoramic radiomorphometric indices and bone mineral density in postmenopausal women. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology. [Internet]. 2010 [cited 2022 Jan 11]; 109(3):[449-56 pp.]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20219601.

171. Dutra V, Devlin H, Susin C, Yang J, Horner K, Fernandes ARC. Mandibular morphological changes in low bone mass edentulous females: evaluation of panoramic radiographs. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology. 2006;102(5):663-8 http://dx.doi.org/10.1016/j.tripleo.2006.02.023

172. Wang L, Cheng XG, Su YB, Brown K, Xu L, Li K, et al. Sex-related variations in cortical and trabecular bone of the femoral neck in an elderly Chinese population. Osteoporosis International. 2017;28(8):2391-9 http://dx.doi.org/10.1007/s00198-017-4043-y

173. Yates LB, Karasik D, Beck TJ, Cupples LA, Kiel DP. Hip structural geometry in old and old-old age: Similarities and differences between men and women. Bone. 2007;41(4):722-32 http://dx.doi.org/10.1016/j.bone.2007.06.001

174. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Preventive Veterinary Medicine. [Internet]. 2000 [cited 2022 February 2]; 45(1):[23-41 pp.]. Available from: https://doi.org/10.1016/S0167-5877(00)00115-X.

175. Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. Canadian Medical Association Journal. 2013;185(11):E537-44 http://dx.doi.org/10.1503/cmaj.121286

176. Leeflang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. Journal of Clinical Epidemiology. 2009;62(1):5-12 http://dx.doi.org/10.1016/j.jclinepi.2008.04.007

177. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. The Lancet. 2005;365(9469):1500-5 http://dx.doi.org/10.1016/S0140-6736(05)66422-7

178. Sundh V, Hange D, Ahlqwist M, Hakeberg M, Lissner L, Jonasson G. FRAX and mandibular sparse trabeculation as fracture predictors: a longitudinal study from 1980 to 2002. European Journal of Oral Sciences. [Internet]. 2017 [cited 2022 Jan 17]; 125(2):[135-40 pp.]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28263008.

179. Bossuyt PM, Reitsma JB, Linnet K, Moons KG. Beyond Diagnostic Accuracy: The Clinical Utility of Diagnostic Tests. Clinical Chemistry. 2012;58(12):1636-43 http://dx.doi.org/10.1373/clinchem.2012.182576

180. Sutthiprapaporn P, Taguchi A, Nakamoto T, Ohtsuka M, Mallick P, Tsuda M, et al. Diagnostic performance of general dental practitioners after lecture in identifying post-menopausal women with low bone mineral density by panoramic radiographs. Dentomaxillofacial Radiology. 2006;35(4):249-52 http://dx.doi.org/10.1259/dmfr/16882209

181. Cootes TF, Taylor CJ, Cooper DH, Graham J. Active shape models-their training and application. Computer vision and image understanding. 1995;61(1):38-59 http://dx.doi.org/10.1006/cviu.1995.1004

182. Aliaga I, Vera V, Vera M, García E, Pedrera M, Pajares G. Automatic computation of mandibular indices in dental panoramic radiographs for early osteoporosis detection. Artificial Intelligence in Medicine. 2020;103:101816 http://dx.doi.org/10.1016/j.artmed.2020.101816

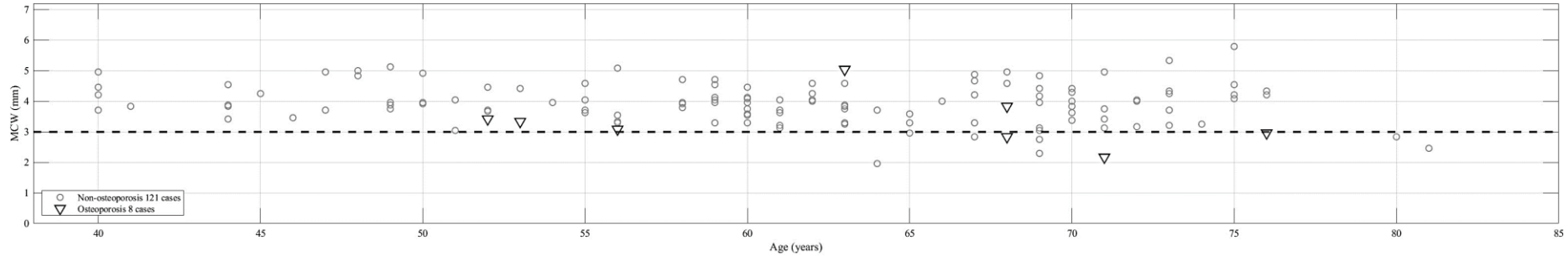
183. Lee K-S, Jung S-K, Ryu J-J, Shin S-W, Choi J. Evaluation of transfer learning with deep convolutional neural networks for screening osteoporosis in dental panoramic radiographs. Journal of clinical medicine. 2020;9(2):392 http://dx.doi.org/10.3390/jcm9020392

184. Åsvold BO, Langhammer A, Rehn TA, Kjelvik G, Grøntvedt TV, Sørgjerd EP, et al. Cohort profile update: the HUNT study, Norway. International Journal of Epidemiology. 2023;52(1):e80-e91 http://dx.doi.org/10.1093/ije/dyac095

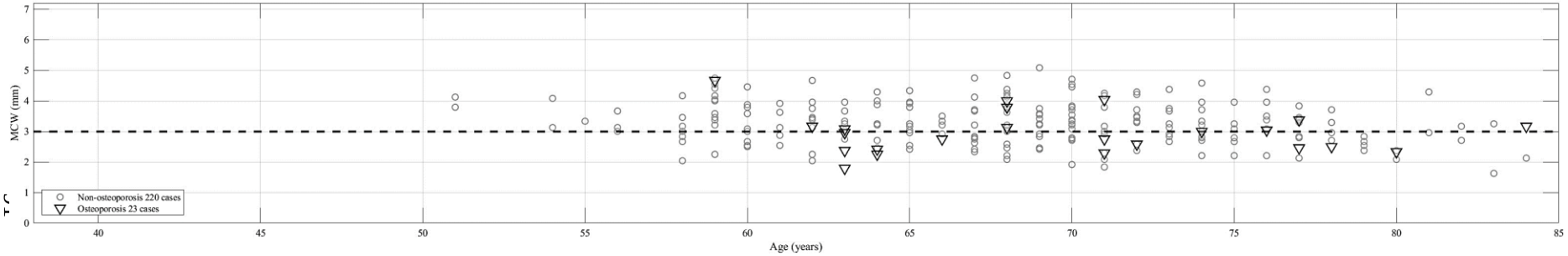
185. Søgaard AJ, Meyer HE, Emaus N, Grimnes G, Gjesdal CG, Forsmo S, et al. Cohort profile: Norwegian epidemiologic osteoporosis studies (NOREPOS). Scandinavian Journal of Public Health. 2014;42(8):804-13 http://dx.doi.org/10.1177/1403494814551858

Appendix I

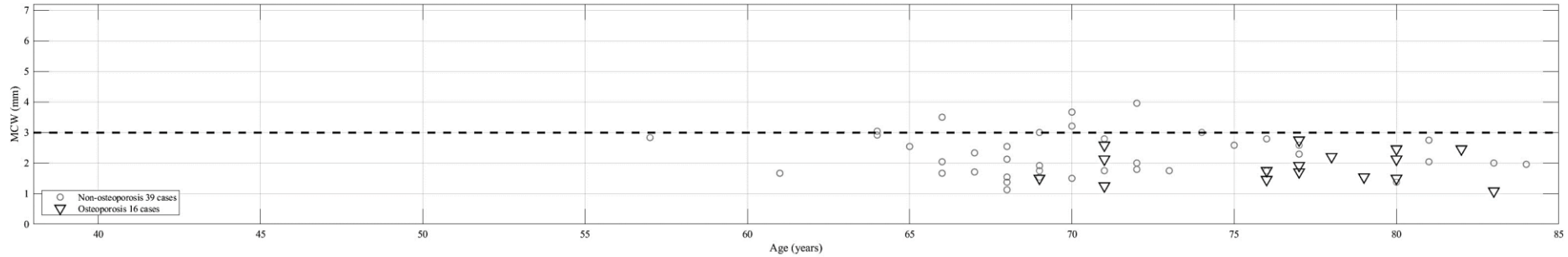
Distribution of osteoporosis (black triangles) and non-osteoporosis cases (grey circles) by age, MCW, and MCI categories in females. The dashed line shows an MCW threshold of 3 mm.


MCW (mm)
Age (years)
Non-osteoporosis 121 cases
Osteoporosis 8 cases

(A)


MCW (mm)
Age (years)
Non-osteoporosis 220 cases
Osteoporosis 23 cases

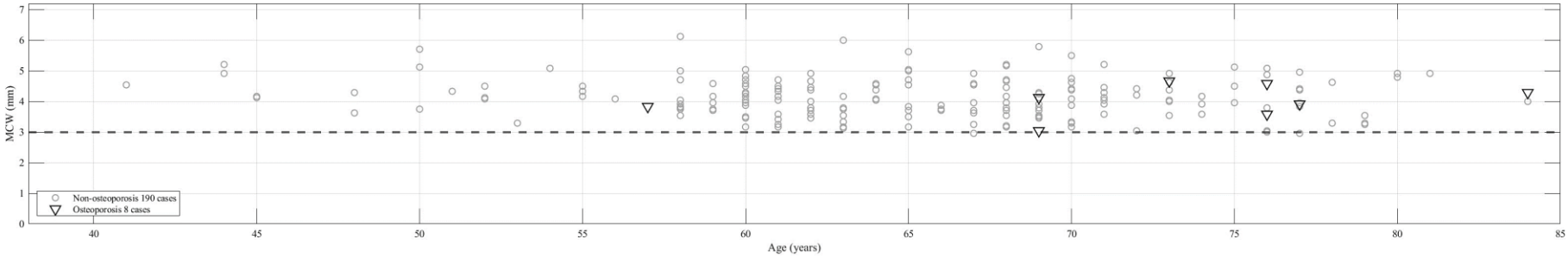
(B)


MCW (mm)
Age (years)
Non-osteoporosis 39 cases
Osteoporosis 16 cases

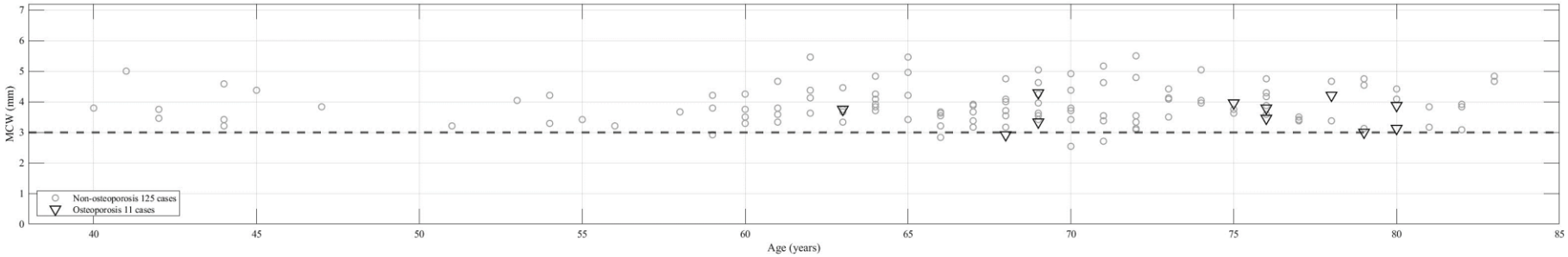
(C)

Appendix II

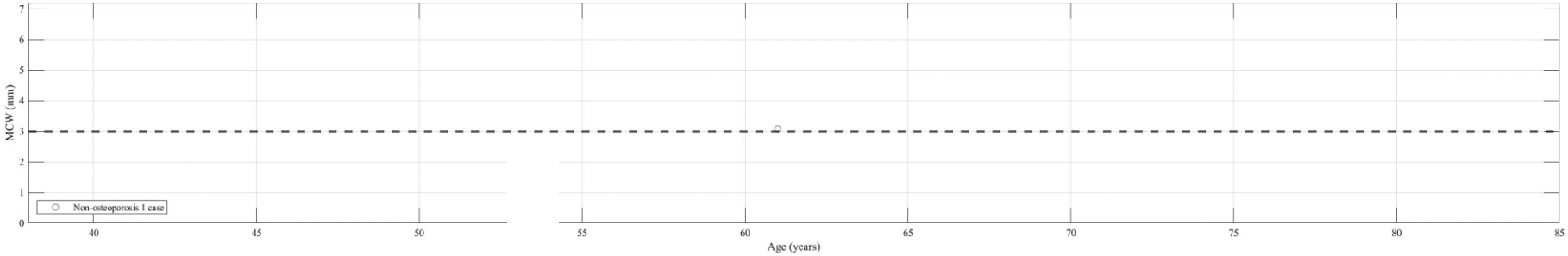
Distribution of osteoporosis (black triangles) and non-osteoporosis cases (grey circles) by age, MCW, and MCI categories in males. The dashed line shows an MCW threshold of 3 mm.


MCW (mm)
Age (years)
Non-osteoporosis 190 cases
Osteoporosis 8 cases

(A)


MCW (mm)
Age (years)
Non-osteoporosis 125 cases
Osteoporosis 11 cases

(B)

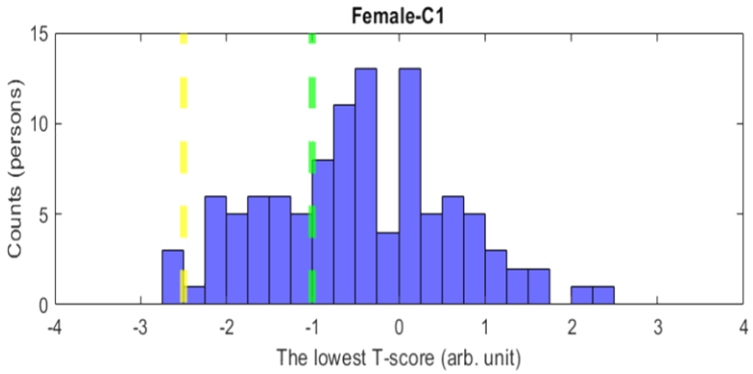

MCW (mm)
Age (years)
Non-osteoporosis 1 case

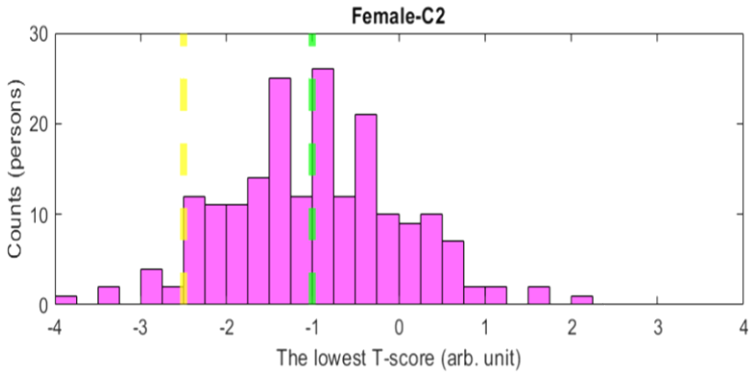
(C)

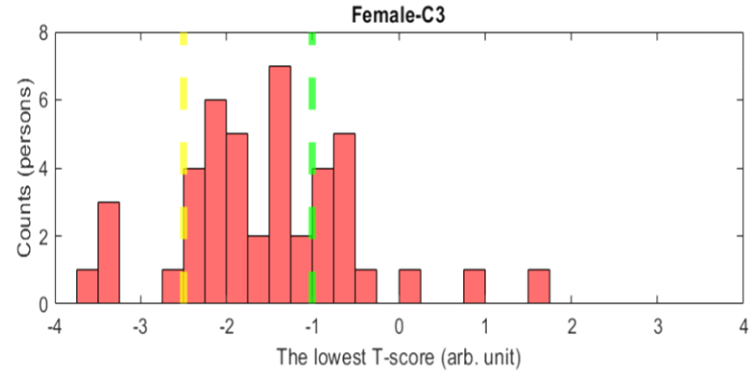
Appendix III

Distribution of T-score in females (A) and males (B) by MCI category.

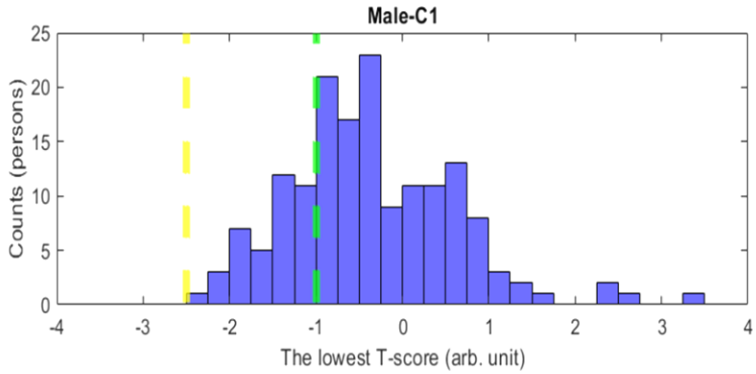
A

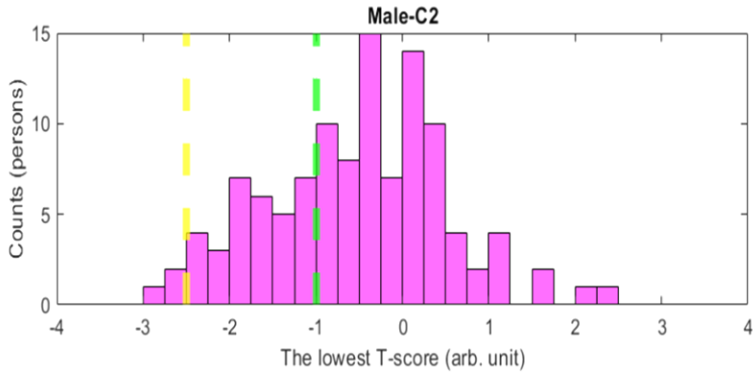

Female-C1
Counts (persons)
The lowest T-score (arb. unit)

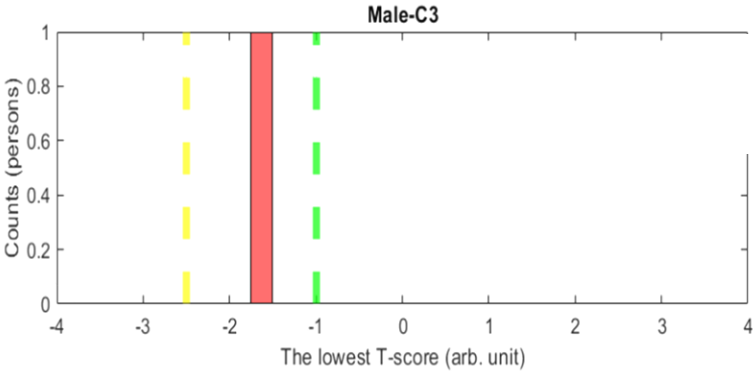

Female-C2
Counts (persons)
The lowest T-score (arb. unit)


Female-C3
Counts (persons)
The lowest T-score (arb. unit)

B


Male-C1
Counts (persons)
The lowest T-score (arb. unit)


Male-C2
Counts (persons)
The lowest T-score (arb. unit)


Male-C3
Counts (persons)
The lowest T-score (arb. unit)

Appendix IV

Diagnostic efficacy estimates for radiomorphometric indices in males.

	Osteoporosis cases (n)	*Non-osteoporosis cases (n)*	*Sensitivity %, 95% CI*	*Specificity%, 95% CI*	*Accuracy %, 95% CI*	*PPV%, 95% CI*	*NPV%, 95% CI*	*+LR, 95% CI*	*-LR, 95% CI*
All males, n=335, Prevalence = 5.67%	***19***	***316***							
MCW≤ 3 mm	2	7	10.5 [1.3; 33.1]	97.8 [95.5; 99.1]	92.8 [89.5; 95.4]	22.2 [6.0; 56.2]	94.8 [93.9; 95.4]	4.7 [1.0; 21.3]	0.9 [0.8; 1.0]
MCW> 3 mm	17	309							
MCI (C3)	0	1	0.0 [0.0; 17.6]	99.7 [98.2; 99.9]	94.0 [90.9; 96.3]				
MCI (C1, C2)	19	315							
MCI (C2, C3)	11	126	57.9 [33.5; 65.6]	60.1 [54.5; 65.6]	60.0 [54.5; 65.3]	8.0 [5.5; 11.6]	95.9 [93.3; 97.6]	1.45 [1.0; 2.8]	0.7 [0.4; 1.2]
MCI (C1)	8	190							
Males, aged ≥ 65, n=200, Prevalence = 8.50%	***17***	***183***							
MCW≤3 mm	2	6	11.8 [1.5; 36.4]	96.7 [93.0; 98.8]	89.5 [84.4; 93.4]	25.0 [6.8; 60.4]	92.2 [90.8; 93.4]	3.6 [0.8; 16.4]	0.9 [0.8; 1.1]
MCW>3 mm	15	177							
MCI (C3)	0	0							
MCI (C1, C2)	17	183							
MCI (C2, C3)	10	81	58.8 [32.9; 81.6]	55.7 [48.2; 63.1]	56.0 [48.8; 62.9]	10.9 [7.4; 15.9]	93.6 [89.1; 96.3]	1.3 [0.9; 2.0]	0.7 [0.4; 1.3]
MCI (C1)	7	102							
MCW or MCI and FRAX®, n=286, Prevalence = 6.29%	***18***	***268***							
MCW≤3 mm and FRAX®>15%	2	13	11.1 [1.4; 34.7]	95.1 [91.8; 97.4]	85.7 [85.8; 93.1]	13.3 [3.6; 38.6]	94.1 [93.1; 94.9]	2.3 [0.6; 9.4]	0.9 [0.8; 1.1]
MCW>3 mm or FRAX®≤15% or both	16	255							
MCI (C3) and FRAX®>15%	0	0							
MCI (C1, C2) or FRAX®≤15% or both	18	268							
MCI (C2, C3) and FRAX®>15%	0	3							
MCI (C1) or FRAX®≤15% or both	18	265							

Paper 1

Teterina, A., Niratisairak, S., Morseth, B. & Bolstad, N. (2023).

Diagnostic efficacy of radiomorphometric indices for predicting osteoporosis in a Norwegian population in the Tromsø Study: Tromsø7.

Oral Surgery, Oral Medicine, Oral Pathology, and Oral Radiology, 135(3), 444-455.

Diagnostic efficacy of radiomorphometric indices for predicting osteoporosis in a Norwegian population in the Tromsø Study: Tromsø7

Anna Teterina, MPH,[a] Sanyalak Niratisairak, PhD,[b] Bente Morseth, PhD,[c] and Napat Bolstad, PhD[a]

Objective. The aim of this study was to investigate the diagnostic efficacy of the radiomorphometric indices of mandibular cortical width (MCW) and mandibular cortical index (MCI) of cortical erosion for osteoporosis screening in adults (≥40 years) and older adults (≥65 years) to determine whether adding a fracture risk assessment tool (FRAX) would improve efficacy.

Study Design. One observer measured MCW and assessed MCI on dental panoramic radiographs acquired for patients in the Tromsø study. These indices, alone and with FRAX scores, were evaluated for efficacy in predicting osteoporosis, which was diagnosed by bone density measurement at the femoral necks with dual-energy X-ray absorptiometry.

Results. MCW ≤3 mm and MCI indicating heavily eroded cortices (C3) had accuracies of 68.8% and 83.6%, respectively, in identifying osteoporosis. In females >65 years, MCW ≤3 mm and C3 produced higher sensitivities but lower specificities, with slightly lower accuracies (61.4% and 79.8%, respectively) compared with all females. The addition of FRAX scores >15% improved the accuracy of MCW ≤3 mm (81.7%) and C3 (87.9%), resulting in high specificity (86.6% and 95.4%). Combining MCW ≤3 mm or C3 with FRAX >15% increased the probabilities of detecting osteoporosis by increasing positive likelihood ratios.

Conclusions. MCW ≤3 mm or MCI C3, when combined with FRAX >15%, showed superior diagnostic efficacy, with high specificity in detecting females without osteoporosis. (Oral Surg Oral Med Oral Pathol Oral Radiol 2023;135:444−455)

Osteoporosis (OP) is a chronic skeletal disease of deteriorating bone microstructure. It can lead to low-trauma bone fractures, which are associated with functional decline, poor quality of life, increased mortality risk, and high health care expenditures.[1] In the United States, the estimated prevalence of OP in females and males aged ≥50 is 19.6% and 4.4%, respectively.[2] The corresponding numbers in the European Union are 21% and 6%, yet some variations by country are observed.[3] A population-based study from Northern Norway revealed 11.5% and 8.3% OP prevalence in females and males.[4]

The reference standard for OP diagnosis is dual-energy X-ray absorptiometry (DEXA), which measures bone mineral density (BMD).[5] The operational definition of the disease is based on the T-score, which is the number of standard deviations of the patient's BMD greater or less than the mean BMD of a 20- to 29-year-old female adult.[1] OP is defined as a T-score less than or equal to −2.5 (i.e., a BMD >2.5 SDs below the mean). However, population-based screening with DEXA and T-scores is not widely practiced in many world regions due to relatively low predictive ability, low availability,[1] and lack of evidence regarding the cost-effectiveness of DEXA.[6]

Instead, various strategies have been adopted in different countries to identify individuals at risk of low-trauma fracture who can benefit from DEXA examination or preventive therapy.[1] The Fracture Risk Assessment Tool (FRAX) is one of those strategies using well-established clinical risk factors. It calculates a percentage value, called the FRAX score, which predicts the 10-year probability of major low-trauma fractures.[7] FRAX includes the assessment threshold score, which is used to refer patients for DEXA examination for refinement of diagnosis, and the intervention threshold score, which is used to start preventive therapy. These thresholds may differ by country.[8]

In Norway, the assessment threshold for FRAX is 15% for major osteoporotic fractures.[9] Individuals ≥65 should particularly be considered for DEXA[10,11] because they sustain hip fractures more frequently, and the overall fracture incidence decreases when this group is treated in a timely manner.[10] However, despite the existing strategy of selecting patients for DEXA, suboptimal examination practice is found in Norway.[12]

Statement of Clinical Relevance

Mandibular cortical width ≤3 mm or severely eroded cortex (C3) assessed on panoramic radiographs, when combined with a fracture risk assessment tool (FRAX score >15%), can be clinically efficacious for osteoporosis screening in females.

[a]Department of Clinical Dentistry, Faculty of Health Sciences, University of Tromsø−The Arctic University of Norway, Tromsø, Norway.
[b]Department of Orthopedics, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway.
[c]School of Sport Sciences, Faculty of Health Sciences, University of Tromsø−The Arctic University of Norway, Tromsø, Norway.
Corresponding author. E-mail address: Napat.l.bolstad@uit.no
Received for publication Apr 8, 2022; returned for revision Sep 22, 2022; accepted for publication Oct 22, 2022.

2212-4403/$-see front matter
https://doi.org/10.1016/j.oooo.2022.10.039

Previous studies have suggested that dental panoramic radiographs (DPRs) are useful for identifying individuals at risk of OP.[13-15] Several radiomorphometric indices based on DPRs have been developed to assess jaw quality and quantity. The most widely used are mandibular cortical width (MCW) and mandibular cortical index (MCI).[16] Two recent systematic reviews with meta-analyses assessed the diagnostic efficacy of MCW and MCI. One review concluded that both indices were useful overall in screening for the low BMD condition of osteopenia,[17] defined as a T-score lower than −1 and greater than −2.5.[17] The other review found that only MCI was useful for detection of osteopenia.[18] However, most of the included studies in both reviews had a substantial risk of bias, specifically in the participant selection domain, which was likely to distort diagnostic efficacy estimates.[17,18] Moreover, the included studies focused mostly on osteopenia,[17] which is not a disease category.[1] Both systematic reviews suggested that further studies should have random participant selection, correct timing, and study flow.[17,18] It is also unclear whether combining MCW and MCI with other case-finding strategies would change their diagnostic efficacy.

To our knowledge, no research has been conducted on the diagnostic efficacy of radiomorphometric indices in Norway. In addition, the diagnostic efficacy estimates from previous studies have limited external validity due to the differences in distribution and severity of OP in various populations,[19] warranting a diagnostic efficacy investigation in the Norwegian population.

Thus, the first aim of the present study was to investigate the diagnostic efficacy of MCW and MCI in the Norwegian adult population. The second aim was to explore how these indices perform in participants aged ≥65. Finally, the study sought to investigate whether FRAX scores combined with MCW or MCI improve diagnostic efficacy.

MATERIALS AND METHODS

Study population

The Tromsø Study is an ongoing population-based study carried out as repeated cross-sectional surveys in Tromsø, a Norwegian city with a population of 71,590. Seven surveys have been performed since 1974.[20] Further information on the survey methods and timeline can be found elsewhere.[20,21] Our study was based mainly on the data from Tromsø7, carried out in 2015 and 2016, and partly on the data from Tromsø6, carried out in 2007 and 2008. In total, 32,591 people living in Tromsø were invited to participate in Tromsø7, and 21,083 people aged 40 to 99 years old consented. They filled out extensive questionnaires on various health-related topics. A total of 3951 randomly selected participants underwent dental clinical examinations, including DPRs. Then, another 3600 participants were randomly selected from the 21,083 consenting patients and underwent an extended general clinical examination, including a BMD examination using DEXA. As a result, 773 participants received both DPRs and DEXA. A flow diagram depicting the selection process is provided in Figure 1.

Ethical considerations

The Tromsø Study was conducted in accordance with the World Medical Association Declaration of Helsinki.[22] The Regional Committee on Research and Ethics and the Norwegian Data Protection Authority (Datatilsynet) approved the Tromsø Study. All participants gave written informed consent. In addition, we received separate approvals from the Regional Committee on Research and Ethics North (reference number 68128) and the Norwegian Centre for Research Data to use the data from the Tromsø Study database.

Panoramic radiography

During the first clinical visit, calibrated dental hygienists acquired DPRs of participants using a Planmeca ProMax 2-Dimensional S3 Dimax-4 panoramic unit (Planmeca Oy, Helsinki, Finland). The exposure parameters were typically set at 68 kV, 8 mA, and 16 seconds but could be changed depending on the patient's size. The receptor was a charge-coupled device with a pixel size of 99 × 99 μm^2 and a digital 16-bit gray level output. The images were stored in tagged image file format, with 101.01 pixels per cm, equivalent to spatial resolution of 5-line pairs per millimeter. No reference object was used to control for magnification; therefore, we adjusted our linear measurements for 20% horizontal and 20% vertical magnifications according to the manufacturer. We excluded the radiographs that were overexposed, underexposed, or unsharp. The MCI reference images described further were also excluded from the analyses. Of 773 DPRs, 762 were eligible for further analyses (Figure 1).

Radiographic measurements of mandibular cortical width

MCW was measured by a single observer (A.T.) on 762 DPRs using ImageJ 1.8.0172 software (US National Institutes of Health, Bethesda, MD, USA).[23] The observer measured MCW bilaterally on the line drawn through the middle of the mental foramen and perpendicular to the inferior border of the mandible as proposed in the original method.[24] If the mental foramen was not visible, the observer measured MCW on the line drawn between the apices of the premolars or approximately in the premolar region when both premolars and the mental foramen were absent on a DPR. The image processing protocol was as follows:

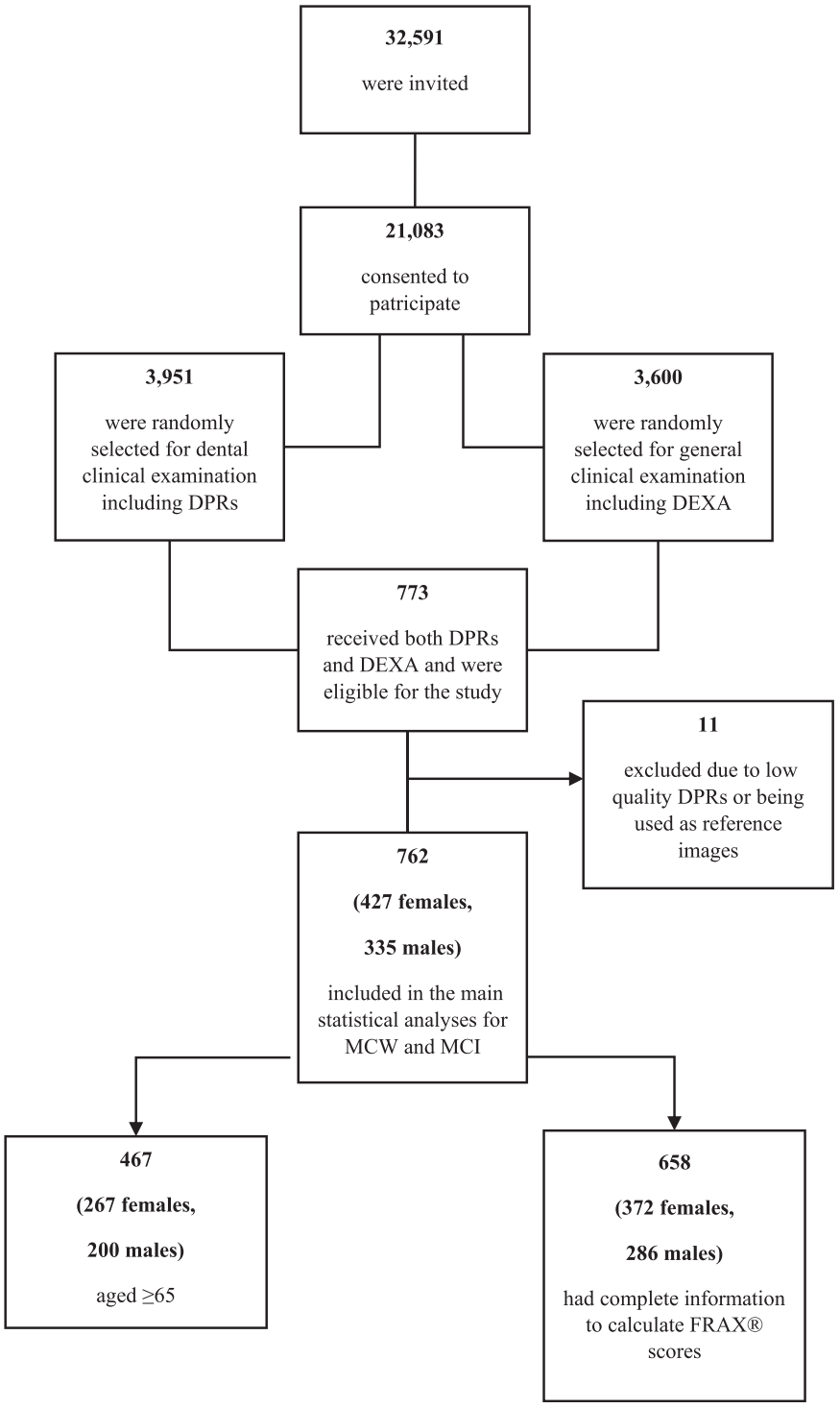


Fig. 1. Flow diagram showing numbers of the study participants. *DPRs*, dental panoramic radiographs; *DEXA*, dual-energy X-ray absorptiometry; *FRAX*, Fracture Risk Assessment Tool.

1. The image was normalized (i.e., the image intensity histogram was extended to utilize the full gray-scale spectrum).
2. The region of interest was manually selected bilaterally in the mental foramen area, including the inferior cortex (Figure 2A). The selected region of interest was subsequently renormalized.
3. A line was drawn through the middle of the mental foramen and perpendicular to the inferior surface of the cortical border of the mandible (Figure 2B).

4. The Gaussian blur filter was applied.
5. A gray-scale plot was made where the lower values represented the darkest area of the image (Figure 2C).
6. The turning points in the gray-scale plot represented the superior (A) and inferior (B) borders of the cortex (Figure 2C). The distance A1B1 corresponded to the thickness of the cortex and was defined as MCW.
7. The average value of right and left MCW represented the MCW of each patient, and 762 MCW values entered by the observer were used for diagnostic efficacy calculations.

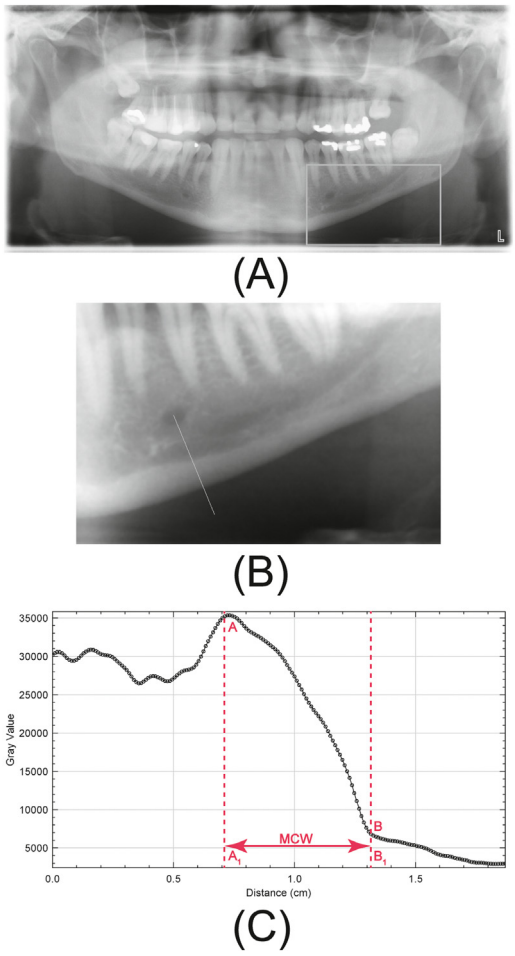


Fig. 2. Mandibular cortical width measured on a grayscale plot in ImageJ. (A). Manual selection of region of interest in the mental foramen area, including the inferior cortex. (B). A line was drawn through the middle of the mental foramen and perpendicular to the inferior border of the mandible. (C). The grayscale plot with turning points representing the superior (A) and inferior (B) borders of the cortex. The distance A1B1 corresponded to the thickness of the cortex and was defined as the mandibular cortical width (MCW).

Mandibular cortical index assessment

Using previously described methods[25] and validated reference images from the study participants, the DPRs were first normalized, and the same observer (A.T.) assessed MCIs bilaterally, distally from the mental foramen. The reference images (Figure 1) were not included in the study. The observer classified mandibular cortical bone as follows: C1, even and sharp endosteal margin on both sides of the mandible (normal cortex); C2, some endosteal cortical residues and/or semilunar defects on 1 or both sides (mildly eroded cortex); and C3, heavy endosteal cortical residues, the cortical bone is clearly porous on 1 or both sides (severely eroded cortex), as depicted in Figure 3.[25] The 762 MCI evaluations made by the observer were used to calculate diagnostic efficacy estimates.

Inter- and Intraobserver Agreements

We carried out a small study on inter- and intraobserver agreements using DPRs from Tromsø7. Two observers, a general dental practitioner with 4 years of experience (A.T.) and an oral radiologist with 20 years of experience (N.B.), participated in a 1-hour training session using the measurement protocol and 20 DPRs randomly selected from the data set of 3951 images from Tromsø7. After the session, they independently assessed the MCW and MCI on 50 randomly selected DPRs from this study's data set of 762 images to calculate interobserver agreement. In addition, 100 DPRs, 50 of which were also used for interobserver agreement calculation, were randomly selected to permit calculation of intraobserver agreement for MCW and MCI. Then, 100 DPRs were assessed twice by the same observer (A.T.) with a 1-month interval between the assessments. All assessments of the MCW and MCI were carried out in a dimmed-light room using the ColorChecker Display Pro device (Calibrite, Wilmington, DE, USA) for screen calibration. The sample sizes of 50 and 100 DPRs were arbitrarily chosen for feasibility reasons.

Clinical risk factors and FRAX score

FRAX scores were calculated using an online tool[26] based on the following well-established clinical risk factors: age, sex, body mass index, previous fracture, a parent with hip fracture, current smoking status, glucocorticoid use, rheumatoid arthritis, secondary OP, and alcohol intake of $\geq$3 units (a unit is defined as a glass of wine or beer or a strong alcohol shot).[26] According to Norwegian guidelines, people having FRAX scores $>$15% should be referred for DEXA examinations.[9,11] At the first visit, the anthropometric parameters of age and sex were recorded, and height and weight were

measured to calculate the body mass index.[21] Data on previous low-trauma fractures was not available in Tromsø7. Therefore, we retrieved this information for the same participants from Tromsø 6, assuming that a negligible number of low-trauma fractures occurred in the study participants between these 2 surveys. Due to the lack of information on a parent with hip fracture in Tromsø7, we used the information on OP in parents from Tromsø 6. The study participants reported smoking status, use of glucocorticoids and other medicines, rheumatoid arthritis, and alcohol intake.[21]

DEXA

DEXA of both hips was carried out during the second clinical visit using a Lunar Prodigy device (GE Healthcare Lunar, Madison, WI, USA). Trained technicians inspected the DEXA scans and made necessary adjustments according to the manufacturer's guidelines. The device was calibrated daily using a phantom. According to the European guidance for the diagnosis and management of OP, we defined osteoporotic cases as the T-score at the femoral neck at least 2.5 SDs below the young female adult mean.[1] We used T-scores at the left and right femoral necks to identify OP. If at least 1 side met the criterion for the disease, the participant was diagnosed with OP. Three participants had unusually high T-scores on 1 side, whereas the T-score on the opposite side was ordinary. In these 3 cases, we used the lower T-score value. The observers were blinded to DEXA measurements, and therefore to the true diagnosis of OP, during the assessment of MCW and MCI.

Statistical analysis

All continuous variables were tested for normal distribution using histograms, the Kolmogorov-Smirnov test, and quantile-quantile plots. The MCW threshold for classification of OP was prespecified to be 3 mm; participants with cortical thickness ≤3 mm were considered to have OP.[27] For MCI, we analyzed diagnostic efficacy at 2 thresholds indicating osteoporosis: heavy cortical erosion (C3) or any cortical erosion (C2, C3). We also combined MCW or MCI and FRAX scores. The study participants were radiographically classified as osteoporotic if they had thin (≤3 mm) cortex, severe cortical erosion (C3) or any cortical erosion (C2, C3), and a FRAX score >15%.

The following diagnostic efficacy parameters were calculated separately for females and males: sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (+LR), and negative likelihood ratio (−LR). First, we calculated diagnostic efficacy for MCW alone (with MCW ≤3 mm vs >3 mm) and MCI alone (with C3 vs C1/C2, and with C2/3 vs C1) in all females (n = 427) and males (n = 335), as indicated in Figure 1. Finally, we created receiver operating characteristic (ROC) curves for MCW by plotting sensitivity against the false positive rate (1 − specificity) at different MCW thresholds.

We then calculated diagnostic efficacy for MCW alone (with MCW ≤3 mm vs >3 mm) and MCI alone (with C3 vs C1/C2, and with C2/3 vs C1) in subgroups of females (n = 267) and males (n = 200) aged ≥65 (Figure 1). We also combined MCW or MCI with FRAX scores (>15% vs ≤15%) to calculate diagnostic efficacy. The total numbers of females and males with complete data on FRAX scores were 372 and 286, respectively (Figure 1).

To define clinically acceptable limits of agreement, we calculated the difference in average MCWs between osteoporotic (2.9 mm) and healthy cases (3.7 mm), which was 0.8 mm. Thus, the calculated values between ±0.8 mm were considered to represent clinically acceptable agreement. To analyze inter- and intraobserver agreements for MCW, Bland-Altman plots were used showing the mean difference and limits of agreement with 95% confidence intervals (CIs). Weighted kappa tests were employed to calculate inter- and intraobserver agreement for MCI classifications.

IBM SPSS Statistics for Windows version 26.0. (IBM SPSS, Armonk, NY, USA), released in 2019, was used to perform analyses of descriptive characteristics, Bland-Altman plots, weighted kappa tests, and ROC curves. The MedCalc diagnostic test evaluation calculator, version 20.014 (MedCalc Software, Ostend, Belgium), was used to calculate sensitivity, specificity, accuracy, PPV, NPV, +LR, −LR, and their CIs.

RESULTS

Because there were only 19 osteoporotic cases in males, the diagnostic efficacy estimates were inconclusive with wide confidence intervals. Therefore, we present the results for males in the supplementary materials (Supplemental Tables S1 and S2, and Supplemental Figure S1 and S2, available at [URL/link*]) and do not discuss them further in the article.

The characteristics of the females are presented in Table I. The average age was 66.6 years. The distribution of T-scores was normal, and the mean was −1.2 SD. Of the 427 females evaluated, 47 had T-scores less than or equal to −2.5 and were therefore diagnosed with OP, for a prevalence of 11.0% of the study population. The MCW values were also normally distributed, with a mean value corrected for magnification of 3.3 mm. More than half of the females (243, or 56.9%) had mildly eroded cortices (C2), and 55 (12.9%) exhibited severe erosion (C3). The distributions of OP by age, MCW, and MCI in females are presented in the

Table I. Characteristics of female participants

Total number	*427*
Mean age, y	66.6
Mean T-score	−1.2 SD
Osteoporosis	47 (11.0%)
Non-osteoporosis	380 (89.0%)
Average MCW corrected for magnification, mm	3.3
MCI C1: Even and sharp endosteal margin (normal cortex)	129 (30.2%)
MCI C2: Some endosteal cortical residues (mildly eroded cortex)	243 (56.9%)
MCI C3: Heavy endosteal cortical residues (severely eroded cortex)	55 (12.9%)

MCW, mandibular cortical width; *MCI*, mandibular cortical index.

supplementary materials (Supplemental Figure S3, available at [URL/link*]).

In the all-female group (Table II), MCW ≤3 mm correctly identified approximately two-thirds of patients with OP (sensitivity of 68.1%) and two-thirds of healthy females (specificity of 69.0%) with an accuracy of 68.8%. In total, 21.3% of females with positive test results (thin cortices) had OP, and 94.6% of females with negative test results (normal cortices) were healthy, as indicated by the PPV and NPV values, respectively. Among the likelihood ratios, +LR was 2.2, −LR was 0.5, and the CIs did not include 1 (i.e., the MCW could distinguish between OP and health).

MCI at the C3 threshold correctly identified 34.0% of osteoporotic and 89.7% of healthy females, producing an accuracy of 83.6%, highest among MCW and MCI criteria. In the females with the positive test result (C3), 29.1% had osteoporosis, as indicated by PPV, which was the highest among other tests. At the same time, NPV of 91.7% was not substantially different from MCW and MCI criteria. The +LR was 3.3, −LR was 0.7, and their CIs did not include the value of 1, indicating MCI (C3) could distinguish between diseased and healthy females. On the contrary, MCI (C2, C3) yielded sensitivity of 83.0% and a specificity of 31.8%. However, the overall accuracy was only 37.5%, PPV was 13.1% compared with 11% prevalence, and the CIs for LRs included the value of 1, indicating that the test could not differentiate osteoporotic and non-osteoporotic cases.

The ROC curve is presented only for MCW, applied to the all-female population (Figure 4). It illustrates the trade-off between sensitivity and the false-positive rate at different thresholds of MCW in females. The area under the ROC curve (AUC) was 0.74, 95% CI (0.67;0.82).

In females aged ≥65 years, in which the prevalence of OP was 13.1%, the sensitivity of each threshold of the radiomorphometric indices was greater than for the population, and the specificity of each criterion was lower. The highest accuracy was still observed for MCI (C3) at 79.8%, but accuracy values for all 3 radiomorphometric thresholds were lower than the corresponding values for the entire population. There were only slight differences in PPVs, NPVs, and LRs compared with the all-female group, indicating almost no changes in diagnostic efficacy but rather a trade-off between sensitivity and specificity (Table II).

The combination of MCW (≤3 mm), MCI (C3) or MCI (C2, C3), and a FRAX score >15% produced several changes in diagnostic efficacy estimates. First, we observed lower sensitivities (30.2%-55.8%) compared with the radiomorphometric indices used alone (34.0%-83.0%). However, specificities were higher, with a range of 72.6% to 95.4%, as opposed to the range of 31.8% to 89.7% with MCW or MCI alone (Table II). Specificity of 95.4% was obtained with the thresholds of MCI (C3) and FRAX >15%. Second, the inclusion of a FRAX score >15% with the MCW or MCI thresholds led to an increase of accuracy values over those obtained with the radiomorphometric thresholds alone, ranging from 70.7% to 87.9%. The highest accuracy, 87.9%, was observed for MCI (C3) and FRAX score >15%. Furthermore, PPVs and +LRs were also notably increased, whereas almost no changes were observed in NPVs and −LRs. Specifically, there was an increase of 17 percentage points in PPV for MCI (C3) when combined with a FRAX score >15% (46.4%) compared with PPV for MCI (C3) alone (29.1%). In addition, +LR for MCI (C3) combined with FRAX score >15% (6.6%) was double the value for MCI (C3) alone (3.3%).

The mean difference in MCW measurements between the 2 observers was −0.11 mm, with upper and lower limits of agreement of 1.3 (1.1; 1.5) mm and −1.5 (−1.7; −1.3) mm, respectively (Figure 5). For repeated measurements by the same observer, the mean difference was 0, the upper limit was 0.8 (0.7;0.9) mm, and the lower limit was 0.7 (−0.8; −0.6) mm (Figure 6). The interobserver agreement (weighted kappa) for MCI was 0.47 [95% CI = (0.30-0.65) (P < .05)], indicating moderate agreement.[28] Intraobserver agreement (weighted kappa) for MCI was 0.72 [95% CI = 0.59-0.85 (P < .05)], indicating substantial agreement.[28]

DISCUSSION

In support of the aim of this study, we found that a thin (≤3 mm) and severely eroded (C3) cortex could differentiate between osteoporotic and healthy females. Our results regarding the diagnostic efficacy of MCW (≤3 mm) and MCI (C3) support previous findings. Regarding MCW, the sensitivity, specificity, and AUC value of 0.74, indicating acceptable discrimination,[29]

Table II. Diagnostic efficacy estimates for the radiomorphometric indices in females

	Osteoporosis cases (n)	*Non-osteoporosis cases (n)*	*Sensitivity %, 95% CI*	*Specificity %, 95% CI*	*Accuracy %, 95% CI*	*PPV %, 95% CI*	*NPV %, 95% CI*	*+LR, 95% CI*	*−LR, 95% CI*
All females, n = 427 Prevalence = 11.0%	47	380							
MCW ≤3 mm	32	118	68.1 (52.8; 80.9)	69.0 (64.0; 73.6)	68.8 (64.2; 73.2)	21.3 (17.5; 25.8)	94.6 (91.9; 96.4)	2.2 (1.7; 2.8)	0.5 (0.3;0.7)
MCW >3 mm	15	262							
MCI (C3)	16	39	34.0 (20.8; 49.3)	89.7 (86.2; 92.6)	83.6 (79.7; 87.0)	29.1 (20.0; 40.3)	91.7 (89.9; 93.1)	3.3 (2.0; 5.4)	0.7 (0.6; 0.9)
MCI (C1, C2)	31	341							
MCI (C2, C3)	39	259	83.0 (69.2; 92.3)	31.8 (27.2; 36.8)	37.5 (32.8; 42.2)	13.1 (11.5; 14.8)	93.8 (88.8; 96.7)	1.2 (1.0; 1.4)	0.5 (0.3; 1.0)
MCI (C1)	8	121							
Females, aged ≥65 n = 267, Prevalence = 13.1%	35	232							
MCW ≤3 mm	27	95	77.1 (59.9; 89.6)	59.0 (52.4; 65.4)	61.4 (55.3; 67.3)	22.1 (18.3; 26.5)	94.5 (90.3; 97.0)	1.9 (1.5; 2.4)	0.4 (0.2; 0.7)
MCW >3 mm	8	137							
MCI (C3)	16	35	45.7 (28.8; 63.3)	84.9 (79.6; 89.3)	79.8 (74.4; 84.4)	31.4 (22.2; 42.3)	91.2 (88.4; 93.4)	3.0 (1.9; 4.9)	0.6 (0.5; 0.9)
MCI (C1, C2)	19	197							
MCI (C2, C3)	31	185	88.6 (73.3; 96.8)	20.3 (15.3; 26.0)	29.2 (23.8; 35.1)	14.3 (12.8; 16.1)	92.2 (81.9; 96.8)	1.1 (1.0; 1.2)	0.6 (0.2; 1.5)
MCI (C1)	4	47							
MCW or MCI and FRAX n = 372, prevalence = 11.5%	43	329							
MCW ≤3 mm and FRAX >15%	19	44	44.2 (29.0; 60.1)	86.6 (82.5; 90.1)	81.7 (77.4; 85.5)	30.2 (21.8; 40.0)	92.2 (90.1; 94.0)	3.3 (2.1; 5.1)	0.6 (0.5; 0.8)
MCW >3 mm or FRAX ≤15% or both	24	285							
MCI (C3) and FRAX >15%	13	15	30.2 (17.2; 46.1)	95.4 (92.6; 97.4)	87.9 (84.1; 92.7)	46.4 (30.7; 62.9)	91.3 (89.6; 92.7)	6.6 (3.4; 13.0)	0.7 (0.6; 0.9)
MCI (C1, C2) or FRAX ≤15% or both	30	314							
MCI (C2, C3) and FRAX >15%	24	90	55.8 (39.9; 70.9)	72.6 (67.5; 77.4)	70.7 (65.8; 75.3)	21.0 (16.2; 26.8)	92.6 (89.9; 94.7)	2.0 (1.5; 2.8)	0.6 (0.4; 0.9)
MCI (C1) or FRAX ≤15% or both	19	239							

Note: Participants were radiographically classified as osteoporotic if they had MCW ≤3 mm, heavy cortical erosion (C3) or any cortical erosion (C2, C3), and a FRAX score >15%.
PPV, positive predictive value; *NPV*, negative predictive value; *+LR*, positive likelihood ratio; *−LR*, negative likelihood ratio; *FRAX*, Fracture Risk Assessment Tool.

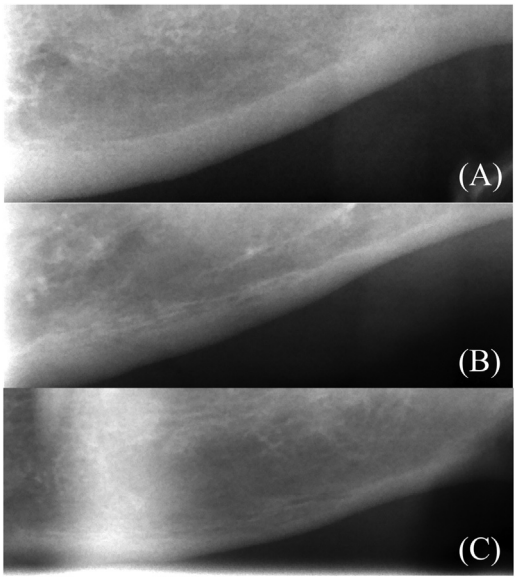


Fig. 3. Reference images used for mandibular cortical index assessment. (A) Even and sharp endosteal margin on both sides; normal cortex (C1). (B) Some endosteal cortical residues and/or semilunar defects on one or both sides; mildly eroded cortex (C2). (C) Heavy endosteal cortical residues; the cortical bone is clearly porous; severely eroded cortex (C3).

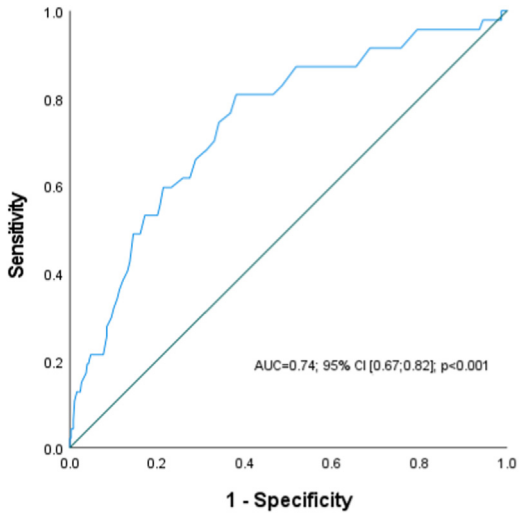


Fig. 4. The receiver operator characteristic (ROC) curve is a plot of the sensitivity of a test against its false-positive rate (1 − specificity). This ROC curve shows how well the mandibular cortical width test performed in females at different thresholds (1.1-5.5 mm). The area under the ROC curve (AUC) with 95% CIs is presented on the plot. The AUC value represents the probability that a randomly selected female with OP will have a thinner cortex, indicating greater suspicion of the disease, than a female without OP. AUC represents the probability that the test will correctly distinguish osteoporosis from health. The *P* value represents the probability that AUC differs from 0.5, a value which would not discriminate osteoporosis from health better than purely by chance. *AUC*, area under the ROC curve.

were consistent with previous studies.[30-32] The AUC value represents the probability that a randomly selected female with OP will have a thinner cortex, indicating greater suspicion of the disease, than a female without OP.[33] The AUC in our investigation was also consistent with AUC values, ranging from 0.65 to 0.76, for various risk−factor-based screening tools for OP currently recommended for clinical use.[34] This finding suggests the clinical importance of MCW. MCI (C3) applied to the all-female group exhibited low sensitivity and high specificity, similar to the results of a recent systematic review.[18]

A diagnostic likelihood ratio (LR) is the percentage of diseased patients with a certain test result, either positive or negative, divided by a percentage of healthy patients with the same test result. The higher the +LR and the lower the −LR, the better a test performs. A likelihood ratio of 1 implies that a test cannot differentiate disease from health. Likelihood ratios can assist the process of making clinical decisions by indicating a post-test probability of disease. In our study, LRs for both MCW (≤3 mm) and MCI (C3) would clinically imply a small change of approximately 15% in the post-test probability of OP.[33] For example, in the Tromsø study, where the prevalence of OP was approximately 11%, we would expect an additional 15% of females (or 26% in total) would have OP after receiving positive test results (i.e., MCW [≤3 mm] and MCI [C3]).

People >65 years are more prone to osteoporotic fracture than younger people.[10] Therefore, our study examined this age group separately from the total population. We found that the diagnostic efficacies of MCW and MCI in females aged ≥65 were generally the same as those in the all-female group, as suggested by their LRs (Table II). However, we observed increased sensitivities and decreased specificities for MCW (≤3 mm) and MCI at both thresholds (C3 and C2, C3) compared with the all-female group. Similar to our study, White et al.[35] found that age >63.5 combined with MCW showed high sensitivity but low specificity. However, their study did not indicate whether the combination of age and MCW had superior diagnostic efficacy compared with MCW alone. The differences in T-score, MCW, and MCI distributions in females aged ≥65 might explain increased sensitivities compared with the all-female group. Similar to another investigation,[36] we observed a tendency for thinner and more eroded cortices with older age (Supplemental Figure S3, available at [URL/link*]). High sensitivity might be clinically preferable in settings where DEXA is proven cost-effective and a health care system covers examination and treatment expenses. To our knowledge, there is no evidence of DEXA cost-effectiveness in Norway.[37]

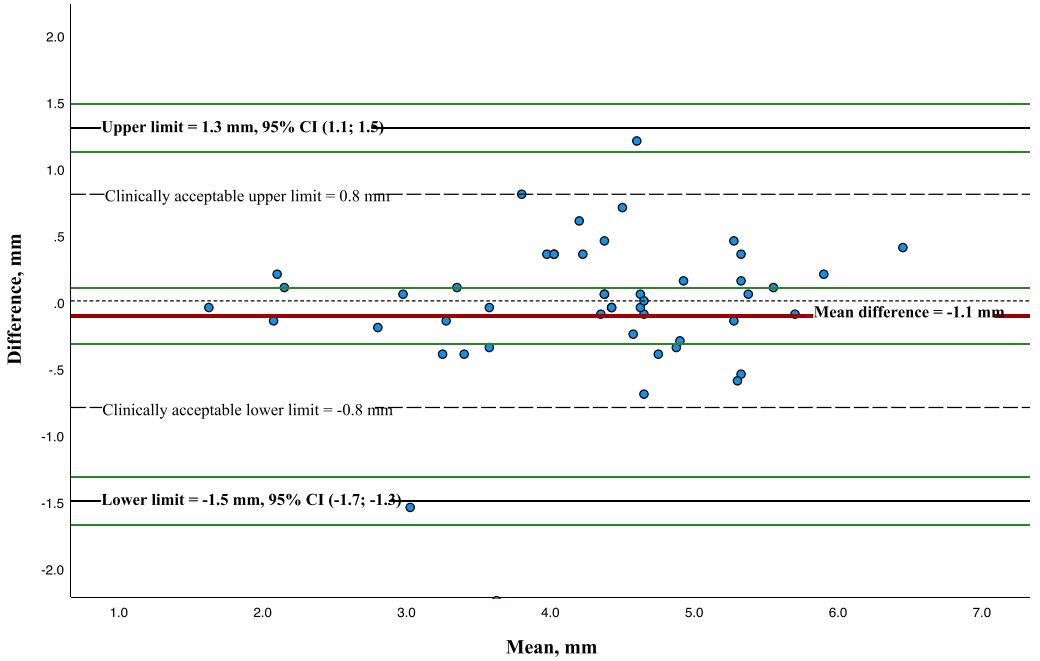


Fig. 5. The Bland-Altman plot shows the mean difference in mandibular cortical width measurements between 2 observers plotted against the mean of the measurements and the limits of agreement with their 95% CIs.

A FRAX score >15% is used as a criterion for referral for the DEXA examination in Norway.[9] FRAX is a convenient online tool that everyone, including dentists, can easily and quickly access. To our knowledge, this was the first study combining radiomorphometric indices with FRAX and using a randomly selected population-based sample. The most important finding of our study was the notable improvement in diagnostic efficacy of the radiomorphometric indices with the addition of a FRAX score >15%, evident from the improvements in +LRs with almost no changes in −LRs (Table II). The highest +LR of 6.6 was found for MCI (C3) combined with FRAX >15%. Clinically, this finding would imply a moderate increase in the post-test probability of OP, revealing 30% to 35% more females with the disease among those with positive test results.[33]

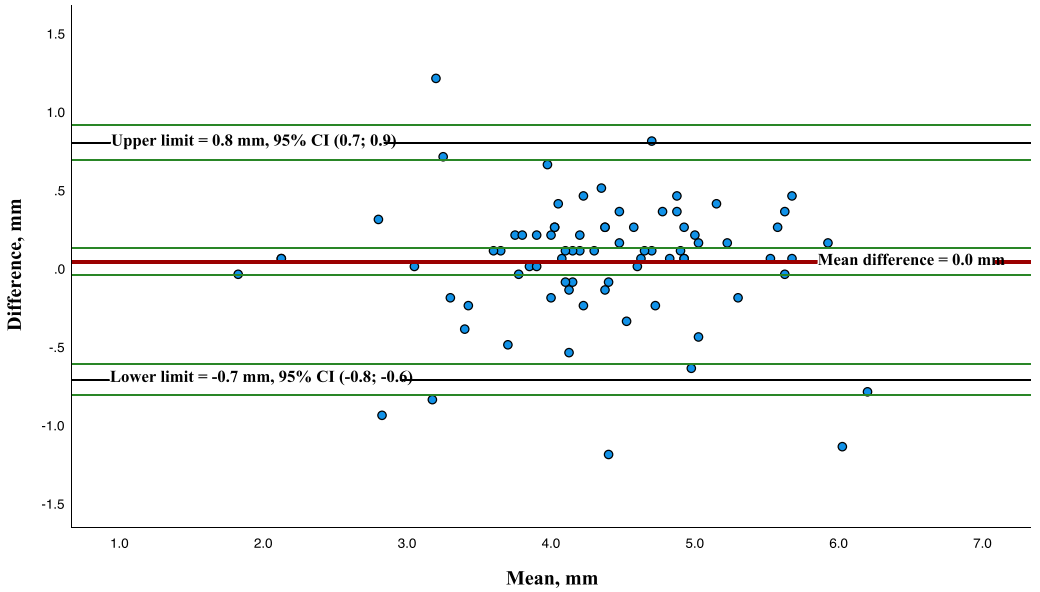


Fig. 6. The Bland-Altman plot shows the mean difference in mandibular cortical width measurements between 2 observations made by the same observer plotted against the mean of the measurements and the limits of agreement with their 95% CIs.

The observed increased diagnostic efficacy for all combinations of radiomorphometric indices with FRAX was, to some extent, consistent with the research on osteoporotic fracture prediction by Sundh et al.[37] In their study, sparse trabeculation on periapical radiographs combined with a FRAX score >15% revealed improved diagnostic efficacy in osteoporotic fracture prediction as shown by improvements in AUCs.[38] In our research, combining FRAX scores with the radiomorphometric indices substantially improved their specificities at the expense of decreased sensitivities. It is a known phenomenon that the sensitivity of 2 tests combined with the "AND" rule (i.e., a disease is assumed present when both tests are positive) is lower than the sensitivities of separate tests, whereas the specificity is higher.[33] We observed a substantial reduction in false-positive results for MCW and MCI combined with FRAX, as indicated by the increased values of specificity and PPV (Table II). This combination of test results might be clinically preferable for OP screening, specifically in settings with fewer than 10.6 DEXA units per million of the general population.[39] Thus, it can apply to Norway, where 44 DEXA units per 5.3 million population were available in 2020.[40]

One may argue that sensitivities, specificities, likelihood ratios, and post-test probabilities of OP were low in our study, suggesting inadequate diagnostic efficacy of the radiomorphometric indices. However, other clinical risk assessment tools for OP screening have presented similar efficacy estimates.[41,42] Nevertheless, these tools are currently recommended for clinical use because effective OP treatment choices exist, with no evident harm from unnecessary screening and treatment.[34] Additionally, predictive values (PPV and NPV) should be interpreted cautiously because they are highly prevalence-dependent.[33] In our study, we used the sample prevalence of OP to calculate predictive values; therefore, they apply only to the sample we used.

The inter- and intraobserver agreement of radiomorphometric indices should also be considered when assessing the usefulness of a screening tool. Our study considered the observers' agreements to be clinically unsatisfactory, specifically considering their CIs. Therefore, future studies should focus on improving the agreement of calculation and interpretation of MCW and MCI. This might be accomplished by the application of artificial intelligence methods to analysis of the mandibular cortex.

Our study had some limitations. The main issue was the small sample size due to low OP prevalence, resulting in wide CIs. Nevertheless, the specificities, NPVs, and AUC for MCW for females were sufficiently precise. The second limitation was the unavoidable error in measurement of MCW due to magnification on DPRs that could have led to a random misclassification of study participants. A previous study suggested using reference objects (metallic ball bearings) to adjust for magnification.[17] However, this plan would be unusual and sometimes unfeasible in dental practice, leaving researchers to accept the imprecision of MCW measurements.

The Tromsø Study has high attendance rates and is considered representative of the population of northern Norway.[20] Thus, the results based on the Tromsø Study data can be generalized at least to this population. Because the Tromsø Study regularly collects extensive clinical information and can be linked to the Norwegian Hip Fracture Register,[43] we recommend further extensive study for more conclusive results on OP screening and fracture prediction using panoramic radiomorphometric indices and FRAX.

Figure 3

CONCLUSIONS

Except for MCI (C2, C3), all indices generally showed some ability to differentiate between osteoporotic and non-osteoporotic females. However, either MCW ≤3 mm or MCI (C3) combined with FRAX >15% yielded superior diagnostic efficacy, especially in detecting females at low risk of osteoporosis.

FUNDING

University of Tromsø−The Arctic University of Norway, Northern Norway Regional Health Authority, the University Hospital of North Norway, and different research funds financed the Tromsø Study. The Department of Clinical Dentistry, Faculty of Health Science (University of Tromsø−The Arctic University of Norway), fully financed the present study.

DECLARATION OF INTEREST

None.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.oooo.2022.10.039.

REFERENCES

1. Kanis JA, Cooper C, Rizzoli R, Reginster J-Y. European guidance for the diagnosis and management of osteoporosis in postmenopausal women. *Osteoporos Int*. 2019;30:3-44.
2. National Center for Health Statistics. Osteoporosis or low bone mass in older adults: United States, 2017-2018. Available at: https://dx.doi.org/ 10.15620/cdc:103477. Accessed June 23, 2022.
3. Hernlund E, Svedbom A, Ivergård M, 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.
4. Emaus N, Omsland TK, Ahmed LA, Grimnes G, Sneve M, Berntsen GK. Bone mineral density at the hip in Norwegian women and men—prevalence of osteoporosis depends on chosen references: the Tromsø Study. *Eur J Epidemiol*. 2009;24:321-328.
5. 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. Swedish Council on Health Technology Assessment (SBU) (SCoHTA). Osteoporosis—prevention, diagnosis and treatment: a systematic review. Available at: Https://www.ncbi.nlm.nih.gov/books/NBK447989/. Accessed June 20, 2022.
7. Kanis JA, Oden A, Johansson H, Borgström F, Ström O, McCloskey E. FRAX and its applications to clinical practice. *Bone*. 2009;44:734-743.
8. Kanis JA, Harvey NC, Cooper C, et al. A systematic review of intervention thresholds based on FRAX. *Arch Osteoporos*. 2016;11:25.
9. Eriksen EF. Osteoporosis. Norwegian electronic handbook for medical doctors. Available at: Https://legehandboka.no/handboken/kliniske-kapitler/endokrinologi/tilstander-og-sykdommer/beinvev-og-vitamin-d/osteoporose. Accessed December 22, 2021.
10. UpToDate. Screening for osteoporosis in postmenopausal women and men. Available at: Https://www.uptodate.com/contents/2046. Accessed Dec 12, 2021.
11. Norwegian Health Institute. Osteoporosis. Norwegian drug handbook. Available at: Https://www.legemiddelhandboka.no/T17.2/Osteoporose. Accessed December 1, 2021.
12. Høiberg M, Rubin K, Gram J, Hermann A, Brixen K, Haugeberg G. Risk factors for osteoporosis and factors related to the use of DXA in Norway. *Arch Osteoporos*. 2015;10:16.
13. Devlin H, Horner K. Mandibular radiomorphometric indices in the diagnosis of reduce skeletal bone mineral density. *Osteoporos Int*. 2002;13:373-378.
14. Taguchi A, Suei Y, Ohtsuka M, Otani K, Tanimoto K, Ohtaki M. Usefulness of panoramic radiography in the diagnosis of postmenopausal osteoporosis in women. Width and morphology of inferior cortex of the mandible. *Dentomaxillofac Radiol*. 1996;25:263-267.
15. Bollen AM, Taguchi A, Hujoel PP, Hollender LG. Case-control study on self-reported osteoporotic fractures and mandibular cortical bone. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. 2000;90:518-524.
16. Yeung AWK, Mozos I. The innovative and sustainable use of dental panoramic radiographs for the detection of osteoporosis. *Int J Environ Res Public Health*. 2020;17:2449.
17. Calciolari E, Donos N, Park JC, Petrie A, Mardas N. Panoramic measures for oral bone mass in detecting osteoporosis: a systematic review and meta-analysis. *J Dent Res*. 2015;94(suppl 3):17S-27S.
18. Kinalski MA, Boscato N, Damian MF. The accuracy of panoramic radiography as a screening of bone mineral density in women: a systematic review. *Dentomaxillofac Radiol*. 2020;49:20190149.
19. Szklo M, Nieto FJ. *Epidemiology: Beyond the Basics*. 4th ed. Sudbury, Ontario, Canada: Jones & Bartlett Learning, LLC; 2018.
20. Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njølstad I. Cohort profile: the Tromso Study. *Int J Epidemiol*. 2012;41:961-967.
21. Hopstock LA, Grimsgaard S, Johansen H, Kanstad K, Wilsgaard T, Eggen AE. The seventh survey of the Tromsø Study (Tromsø7) 2015-2016: study design, data collection, attendance, and prevalence of risk factors and disease in a multipurpose population-based health survey. *Scand J Public Health*. 2022;50:919-929.
22. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;310:2191-2194.
23. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012;9:671-675.
24. Ledgerton D, Horner K, Devlin H, Worthington H. Panoramic mandibular index as a radiomorphometric tool: an assessment of precision. *Dentomaxillofac Radiol*. 1997;26:95-100.
25. Klemetti E, Kolmakov S, Kröger H. Pantomography in assessment of the osteoporosis risk group. *Scand J Dent Res*. 1994;102:68-72.
26. Centre for Metabolic Bone Diseases, University of Sheffield, UK. FRAX Fracture Risk Assessment Tool. Available at: https://www.sheffield.ac.uk/FRAX/index.aspx. Accessed December 22, 2021.
27. Taguchi A, Tanaka R, Kakimoto N, et al. Clinical guidelines for the application of panoramic radiographs in screening for osteoporosis. *Oral Radiol*. 2021;37:189-208.
28. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
29. Hosmer DW Jr, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. 3rd ed. New York, NY: John Wiley & Sons; 2013.
30. Balto KA, Gomaa MM, Feteih RM, et al. Dental panoramic radiographic indices as a predictor of osteoporosis in postmenopausal Saudi women. *J Bone Metab*. 2018;25:165-173.
31. Leite AF, Figueiredo PT, Guia CM, Melo NS, de Paula AP. Correlations between seven panoramic radiomorphometric indices and bone mineral density in postmenopausal women. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. 2010;109:449-456.
32. Karayianni K, Horner K, Mitsea A, et al. Accuracy in osteoporosis diagnosis of a combination of mandibular cortical width measurement on dental panoramic radiographs and a clinical risk index (OSIRIS): the OSTEODENT project. *Bone*. 2007;40:223-229.
33. Zhou X-H, McClish DK, Obuchowski NA. *Statistical Methods in Diagnostic Medicine*. Hoboken, NJ: John Wiley & Sons; 2011.
34. Viswanathan M, Reddy S, Berkman N, et al. Screening to prevent osteoporotic fractures: updated evidence report and systematic review for the US Preventive Services Task Force. *JAMA*. 2018;319:2532-2551.
35. White SC, Taguchi A, Kao D, et al. Clinical and panoramic predictors of femur bone mineral density. *Osteoporos Int*. 2005;16:339-346.
36. Mizukuchi T, Naitoh M, Hishikawa T, et al. Automatic measurement of mandibular cortical bone width on cone-beam computed tomography images. *Oral Radiol*. 2021;37:412-420.
37. Joakimsen RM, Fønnebø V. Screening for osteoporosis in Norway? *Tidsskr Nor Laegeforen*. 2001;121:615-618. [in Norwegian].
38. Sundh V, Hange D, Ahlqwist M, Hakeberg M, Lissner L, Jonasson G. FRAX and mandibular sparse trabeculation as fracture predictors: a longitudinal study from 1980 to 2002. *Eur J Oral Sci*. 2017;125:135-140.
39. Ström O, Borgström F, Kanis JA, et al. Osteoporosis: burden, health care provision and opportunities in the EU. *Arch Osteoporos*. 2011;6:59-155.
40. Norwegian Osteoporosis Association. Overview of places in Norway where bone mineral density examination is available.

Available at: https://nof-norge.no/wp-content/uploads/2020/02/dexa_maalesteder_april_2020.pdf. Accesssed June 22, 2022.

41. Crandall CJ, Larson J, Gourlay ML, et al. Osteoporosis screening in postmenopausal women 50 to 64 years old: comparison of US preventive services task force strategy and two traditional strategies in the Women's Health Initiative. *J Bone Miner Res*. 2014;29:1661-1666.
42. Pang WY, Inderjeeth CA. FRAX without bone mineral density versus osteoporosis self-assessment screening tool as predictors of osteoporosis in primary screening of individuals aged 70 and older. *J Am Geriatr Soc*. 2014;62:442-446.
43. Eggen AE, Mathiesen EB, Wilsgaard T, Jacobsen BK, Njølstad I. The sixth survey of the Tromso Study (Tromse 6) in 2007-08: collaborative research in the interface between clinical medicine and epidemiology: study objectives, design, data collection procedures, and attendance in a multipurpose population-based health survey. *Scand J Public Health*. 2013;41:65-80.

Paper 2

Teterina, A., Niratisairak, S., Morseth, B. & Bolstad, N. (2023).

General and local predictors of mandibular cortical bone morphology in adult females and males: The seventh survey of the Tromsø Study.

Clinical Oral Investigations, 27, 6577-6587.

Clinical Oral Investigations
https://doi.org/10.1007/s00784-023-05263-0

RESEARCH

General and local predictors of mandibular cortical bone morphology in adult females and males: the seventh survey of the Tromsø Study

Anna Teterina[1] · Sanyalak Niratisairak[2,3] · Bente Morseth[4] · Napat Bolstad[1]

Received: 2 June 2023 / Accepted: 11 September 2023

Abstract

Objectives To analyze factors predicting mandibular cortical width (MCW) and mandibular cortical index (MCI) in adult females and males.

Material and methods Data on 427 females and 335 males aged 40–84 from The Tromsø study: Tromsø7 were used. *T*-score, age, menopausal status (for females), remaining teeth, and periodontal status were analyzed in linear and logistic regression analyses as predictors of MCW and MCI, respectively.

Results *T*-score, age, and the number of remaining teeth significantly predicted MCW in females but not males. Standardized β coefficients were 0.286, −0.231, and 0.131, respectively. The linear regression model explained 24% of MCW variation in females. MCI in females was significantly predicted by *T*-score, age, and remaining teeth with the Wald values of 9.65, 6.17, and 5.83, respectively. The logistic regression model explained 16.3−23% of the variation in MCI in females. In males, *T*-score was the only significant predictor of the eroded cortex, and the logistic model explained only 4.3–5.8% of the variation in MCI.

Conclusions The *T*-score demonstrated a stronger relationship with MCW and MCI than other factors in females, which supports the usefulness of those indices for osteoporosis screening. Conversely, the *T*-score exhibited no association with MCW and remained the only significant predictor of MCI in males, yet to a lesser extent than in females.

Clinical relevance Understanding factors affecting mandibular cortical morphology is essential for further investigations of MCW and MCI usefulness for osteoporosis screening in females and males.

Keywords Mandibular cortical width · Mandibular cortical index · Panoramic radiographs · Osteoporosis · Dual-energy x-ray absorptiometry

Introduction

Osteoporosis is a chronic, non-communicable disease that deteriorates bone tissue and makes bones fragile and prone to fractures which are considered a public health problem due to increased mortality risk and considerable health costs [1–3]. Genetic factors and sex define bone mass and structure to a great extent, and osteoporosis is more prevalent in females than in males [3, 4]. Both males and females reach their peak bone mass approximately in their 20s; after that, gradual bone loss starts in both sexes in their third decade due to reduced osteoblast activity [3, 5], and accelerates faster in females in post-menopause due to declining estrogen levels [6].

There are associations between jawbone morphology and the state of bone tissue in the whole body. For example, a moderate positive correlation was found between the

✉ Anna Teterina
anna.teterina@uit.no

1 Department of Clinical Dentistry, Faculty of Health Sciences, University of Tromsø–The Arctic University of Norway, Tromsø, Norway

2 Department of Orthopaedics, Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

3 Biomechanics Laboratory, Rikshospitalet, Division of Orthopaedic Surgery, Oslo University Hospital, Oslo, Norway

4 School of Sport Sciences, Faculty of Health Sciences, University of Tromsø–The Arctic University of Norway, Tromsø, Norway

Published online: 22 September 2023

mandible's bone mineral density (BMD), specifically BMD of the buccal mandibular cortex, and femoral neck BMD [7]. Furthermore, the morphology of the mandibular cortex assessed on dental panoramic radiographs was associated with skeletal bone turnover in elderly females [8]. These findings were supported by further research collected in two systematic reviews showing that changes in mandibular cortical morphology on dental panoramic radiographs could predict low BMD or osteoporosis in women [9–11]. Two radiomorphometric indices, i.e., mandibular cortical width (MCW) and mandibular cortical index (MCI), were extensively studied as potentially useful for osteoporosis screening [12].

Nevertheless, little is known about the extent to which different factors are associated with the morphology of the mandibular cortex, while this knowledge is essential for supporting or arguing against using radiomorphometric indices for osteoporosis screening. Such studies are specifically lacking in males [13, 14]. Some studies have shown that age and gender are significant predictors of the thin and eroded cortex; however, their analyses did not account for T-score or other possible factors [15, 16].

Mechanical loading is also important for building and maintaining bone tissue [17]. An animal study showed that rats with a soft diet had lower BMD of mandibles than rats eating solid food [18]. Dental practitioners observe alveolar bone loss after tooth extraction, although it occurs to various extents in different individuals [17]. Tooth loss may affect parts of the mandible other than the alveolar ridge; Taguchi et al. found that the number of remaining teeth adjusted for age was related to mandibular cortex morphology in women [19]. Dutra et al. found an association between remaining teeth and cortical thickness in the antegonial region of the mandible, which might confound relationships between cortical thickness and osteoporosis [20].

One of the most common reasons for tooth loss is periodontitis — an inflammatory disease affecting tooth-supportive structures called periodontium. Dental plaque bacteria induce inflammation in the periodontium, which is subsequently modified by a host immune response. Inflammatory cells release cytokines that activate osteoclasts, while the latter initiate resorption of the alveolar bone surrounding teeth [21]. The mechanistic links between osteoporosis and periodontitis have been studied but remained unclear [22]. There is emerging evidence that patients with periodontitis exhibit a general inflammatory response with elevated levels of C-reactive protein and inflammatory cytokines, including those activating osteoclasts, while osteoclasts are responsible for bone resorption [21, 23–25]. Thus, periodontitis might also be a factor influencing the mandibular cortical bone.

To our knowledge, none of the previous studies have examined the contribution of different factors to the morphology of the mandibular cortex. This study aimed to analyze the relationship between general factors such as skeletal BMD, age, and menopausal status (for females), local factors, such as the number of teeth and periodontal status, and the morphology of mandibular cortex measured by MCW and MCI on dental panoramic radiographs in males and females.

Material and methods

The Tromsø Study is an ongoing population-based study initiated in 1974 and carried out as repeated cross-sectional surveys in Tromsø, Norway. The data from the seventh survey (Tromsø7) was used in the current study. All inhabitants of Tromsø aged over 40 (n=32,591) were invited to participate in Tromsø7, and 21,083 consented, yielding a response rate of 65% [26]. They filled out extensive questionnaires on various health-related topics. All study participants reported their ages and sex, and females reported their menopausal status in the questionnaires. Further information on the Tromsø7 sampling procedure and data collection is available elsewhere [26].

Three thousand nine hundred fifty-one randomly selected participants (Fig. 1) underwent a dental panoramic radiograph (DPR) examination and extensive dental clinical examination during the first visit. Planmeca ProMax 2-Dimensional S3 Dimax-4 panoramic unit (Planmeca Oy, Helsinki, Finland) was used for DPRs acquisition. During a dental examination, bleeding on probing (BOP) and periodontal probing depth (PPD) were recorded at four sites of each toot at all teeth except the third molars. PPD was measured to the closest millimeter with a periodontal probe (WHO probe LM555B). Radiographic marginal bone level (RBL) of interproximal surfaces of all teeth except third molars was measured on DPRs, according to Holde et al. [27]. Periodontitis was defined according to the classification by the American Academy of Periodontology and European Federation of Periodontology, launched in 2017 [28]. Prevalent periodontitis was determined and further classified by stages if RBL was observed at two or more non-adjacent teeth and further classified by stages. Stage I was defined as RBL<15% and PPD ≤4 mm, stage II as RBL of 15–33% and PPD ≤5 mm, or RBL<15% and PPD 5 mm, and stage III–IV as RBL>33%, or <33% RBL and PPD ≥6 mm. Stage III and IV collapsed due to a few cases. Information on the reason for tooth loss, complexity factors, vertical bone loss, furcation involvement, ridge defects, tooth mobility, masticatory dysfunction, and bite collapse/drifting/flaring was unavailable. Non-periodontitis cases were defined as BOP at less than 10% of sites, and gingivitis cases as BOP ≥10% [29]. Periodontal stability was defined as RBL detectable at two or more non-adjacent teeth but no PPD >3 mm.

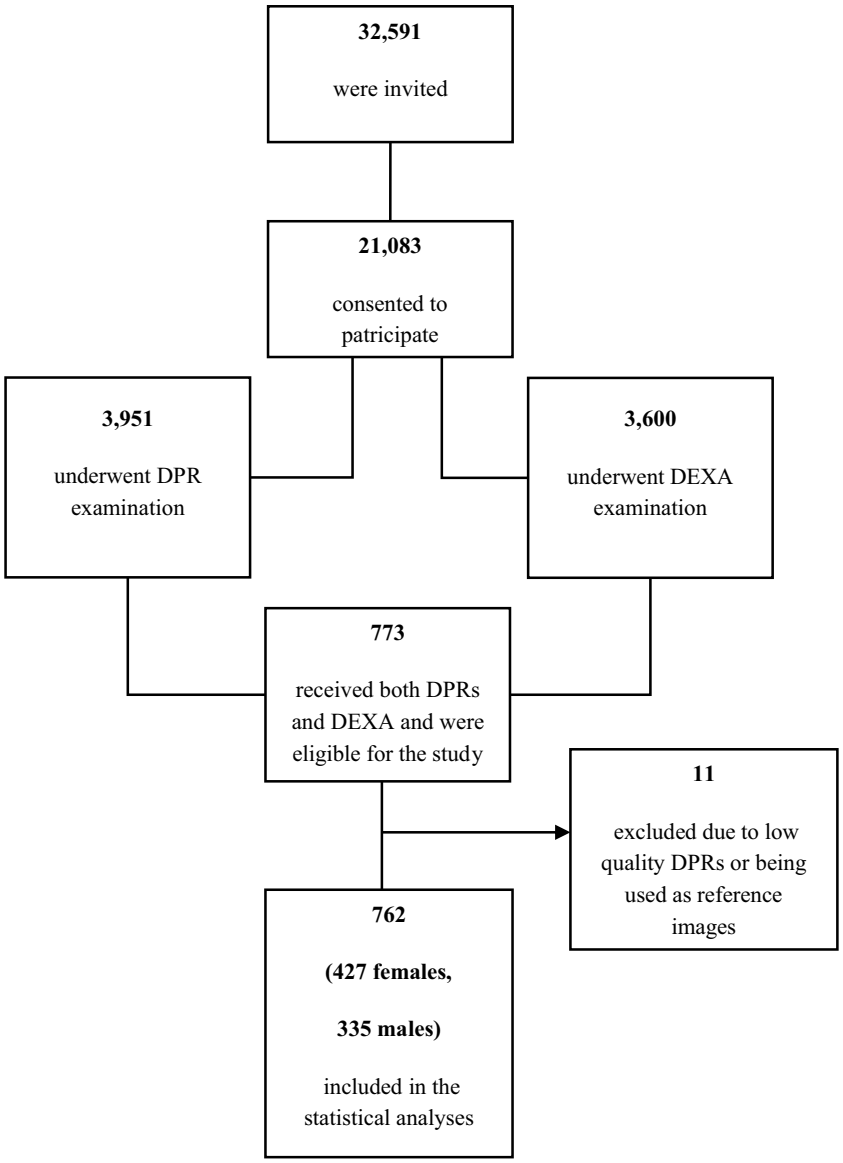


Fig. 1 Selection of the study participants

At the second visit, another 3600 participants, who were randomly selected from the 21,083 Tromsø7 participants attending the first visit, received bone mineral density examination (BMD) using dual-energy x-ray absorptiometry (DEXA) (Fig. 1). Lunar Prodigy device (GE Health-care Lunar, Madison, WI, USA) was used to measure BMD and *T*-scores at the left and right femoral necks. *T*-score is a ratio of the difference between the patient's BMD and the mean BMD of the young sex-matched adult (reference BMD) to the standard deviation of the reference BMD. This study used a Lunar reference from the US reference population. The *T*-score was expressed in standard deviations (S.D.). The minimum *T*-score from either the left or right femoral neck of each participant was used in this study.

Seven hundred seventy-three participants aged 40 to 84 underwent DEXA and DPR examinations and were included in our cross-sectional study (Fig. 1). Eleven DPRs were excluded due to inferior quality or being used as reference images for MCI assessment. We measured MCW and MCI using ImageJ 1.8.0172 software (U.S. National Institutes of Health, Bethesda, MD, USA) [30] on the rest 762 DPRs. MCW was measured bilaterally along the line drawn through the middle of the mental foramen and perpendicular to the lower mandibular border, as proposed by Ledgerton et al. [31], where MCW is the shortest distance between the upper and the lower borders. In this study, we use an average MCW value of the left and the right side of individuals, and the unit of MCW is a millimeter. Since no reference object was used to control for magnification, all linear measurements were adjusted for a magnification factor of 20%, indicated by the manufacturer. MCI of the cortex was classified into three categories: C1, even and sharp endosteal margin

on both sides of the mandible (normal cortex); C2, some endosteal cortical residues, and semilunar defects on one or both sides (mildly eroded cortex); and C3, heavy endosteal cortical residues, the cortical bone is porous on one or both sides (severely eroded cortex). Details of image processing, MCW, and MCI measurements are available from our previous study [32].

Statistical analyses were conducted using IBM SPSS Statistics for Windows version 26.0. MATLAB (version R2021b, The MathWorks Inc., Natick, Massachusetts) was used to make plots. The normality of distribution was assessed by visual examinations of histograms, Q-Q plots, skewness, and kurtosis values. All continuous predictors were normally distributed except the number of remaining teeth. Inter- and intra-observer reliabilities of MCW and MCI were reported in the previous study [32].

Person and Spearman's correlation coefficients were calculated for normally and non-normally distributed continuous predictors to analyze their correlations with MCW. Hierarchical linear regression analysis was used to analyze relationships among general predictors (T-score, age, and menopausal status), local predictors (remaining teeth and periodontal status), and MCW. Age was added to the first block, *T*-score, and menopausal status were added to the second block, while remaining teeth and periodontal status were added to the third block. Standardized β coefficients were used to compare the strength of associations between the predictors and MCW. The linear model assumptions were met, and influential cases were detected in neither females nor males.

Hierarchical logistic regression analysis with the same blocks was used to assess relationships among general, oral predictors, and MCI. MCI was used as a binary outcome (C1 – even and smooth cortex vs. C2, C3 – mildly or severely eroded cortex) due to the low number of participants having the C3 category (55 females and one male). Wald statistics were used to compare the strength of associations between the predictors and MCI. The logistic model assumptions were met.

All regression analyses were carried out separately for males and females. Age was used as a continuous predictor in linear regression and a categorical predictor in logistic regression with the following groups: 40–49; 50–59; 60–69; 70–79; 80+. *T*-score was used as a continuous predictor with a 0.1 SD increment. Periodontal status was divided into three following groups: non-periodontitis or mild periodontitis (health, gingivitis, stage I), moderate periodontitis (stage II), or severe periodontitis (stage III, IV) in both linear and logistic regression analyses. Missing values were excluded pairwise. Data on menopausal status was available for 419 females. Periodontal status for edentulous individuals and individuals with periodontal stability was not included in the regression model, while there were 15 females and 6 males for whom periodontal data was missing. Thus, the total number of females and males having data on periodontal status and included in the regression analysis was 310 and 272, respectively. Other predictors did not have missing values.

Results

Table 1 presents the characteristics of the study participants by sex. The average MCW was 3.3 mm in females and 4.0 mm in males. Most females (57%) had mildly eroded cortexes, 12.9% had severely eroded, and 30.1% had dense cortexes. Most males (59.1%) had dense cortexes, 40.6% had mildly eroded cortexes, and only one male had severely eroded cortex.

All continuous predictors were significantly correlated with MCW in females (Table 2). Age showed a negative correlation with MCW, with a coefficient of −0.38. Figure 2a presents the means and the distributions of MCW in females by age, showing a tendency to have a thinner cortex with age. *T*-score and the number of remaining teeth showed positive correlations with MCW with coefficients of 0.40 and 0.34, respectively (Table 2). Figure 2b shows a tendency for a thinner cortex with a decreasing *T*-score. In the simple linear regression analyses in females, all predictors except periodontal status were significantly associated with MCW,

Table 1 Characteristics of the study participants

	Females, *n*=427	Males, *n*=335
Mean MCW, mm, (S.D.)	3.3 (± 0.7)	4.0 (±0.6)
MCI		
C1 (dense cortex)	129 (30.1%)	198 (59.1%)
C2 (mildly eroded cortex)	243 (57.0%)	136 (40.6%)
C3 (severely eroded cortex)	55 (12.9%)	1 (0.3%)
Mean age, years, (SD)	66.6 (±8.6)	66.2 (±8.8)
Mean T-score, S.D.s, (SD)	-1.3 (±1.0)	-1.0 (±0.9)
Menopausal status*		
Premenopausal	23 (5.5%)	-
Postmenopausal	396 (94.5%)	-
The median number of remaining teeth	24	24
Edentulous	25 (5.9%)	18 (5.4%)
One or more teeth	327 (76.5%)	258 (77.0%)
Full dentition	75 (17.6%)	59 (17.6%)
Periodontal status**:		
Healthy periodontium, gingivitis or periodontitis, stage I[b]	72 (23.2%)	43 (15.8%)
Periodontitis, stage II[b]	162 (52.3%)	136 (50.0%)
Periodontitis, stage III–IV[b]	76 (24.5%)	(34.2%)

*Number of observations=419

**Number of observations in females=310; in males=272

Table 2 The results of simple (unadjusted β) and multiple (adjusted β) linear regression analysis predicting MCW in females and males

Predictors	r	Unadjusted β, 95% CI	S.E.	Adjusted β, 95% CI	S.E.	Standardized β	Characteristics of the model
Females							
T-score	0.40**	0.35 (0.27; 0.43) **	0.0	0.25 (0.16; 0.35) **	0.05	0.286 **	Constant =0.47 $F = 18.7$ $p < 0.001$ $R^2 = 0.240$ R^2 adjusted = 0.227
Age	−0.38**	−0.04 (−0.05; −0.03) **	0.00	−0.02 (−0.04; −0.01) **	0.01	−0.231**	
Menopausal status		−0.70 (−0.10; −0.34) **	0.18	0.11 (−0.34; 0.56)	0.23	0.029	
Remaining teeth	0.34**	0.03 (0.02; 0.04) **	0.00	0.01 (0.001; 0.03) *	0.01	0.131*	
Periodontal status		0.00 (−0.14;14)	0.07	0.03 (−0.09; 0.16)	0.06	0.025	
Males							
T-score	0.14*	0.09 (0.02; 0.16) *	0.03	0.08 (0.00; 0.16)	0.04	0.123	Constant = 0.39 $F = 2.3$ $p = 0.059$ $R^2 = 0.033$ R^2 adjusted = 0.019
Age	−0.07	0.00 (−0.01; 0.00)	0.00	0.00 (−0.01; 0.01)	0.01	0.023	
Remaining teeth	0.10	0.01 (0.00; 0.02)	0.00	0.01 (0.00; 0.02)	0.01	0.116	
Periodontal status		0.05 (−0.12; 0.09)	0.05	−0.03(−0.14; 0.0)	0.06	−0. 036	

*Results are significant at 0.05 level

**Results are significant at 0.001 level

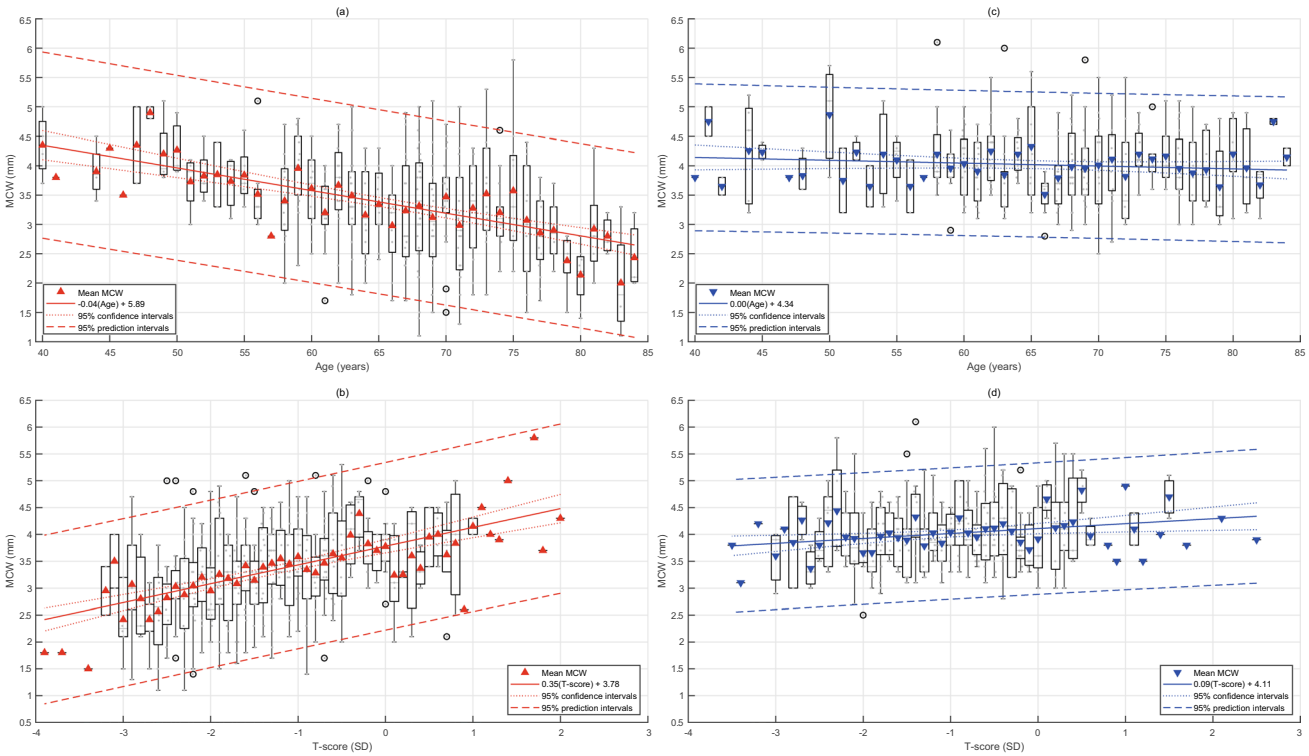


Fig. 2 Box plots show mean MCW (filled triangles), its confidence intervals (boxes), and its distribution (vertical lines) in different ages in females (**a**) and males (**c**) and for different T-scores in females (**b**) and males (**d**). Solid lines are regression lines, while dotted lines are confidence intervals for the regression lines

as suggested by unadjusted β coefficients with their confidence intervals (Table 2). In multiple linear regression analysis, *T*-score, age, and the number of teeth remained significantly associated with MCW. Every 0.1 SD lowered *T*-score resulted in 0.25-mm thinner MCW. One year increase in age was associated with a 0.02-mm reduction in MCW in females. Each additional remaining tooth was associated with having a 0.01-mm thicker cortex. *T*-score contributed the most to the variation in MCW in females, with the standardized β of 0.286, followed by age and remaining teeth with the standardized β's of −0.231 and 0.131, respectively. The multiple linear regression model for females significantly

predicted MCW (F=18.7, p<0.001). The model explained 24.0% of the variation in MCW in females (R^2=0.240).

In males, only T-score correlated with MCW (r=0.14, p=0.009) (Table 2). Males did not tend to have thinner cortexes with older age (Fig. 2c). A significant association between the T-score and MCW with an unadjusted β coefficient of 0.09 was observed in males in simple linear regression analysis. Figure 2d shows that males tended to have a thinner cortex with decreasing T-score. However, this tendency was less pronounced in males than females. In the multiple linear regression analysis, none of the predictors showed associations with MCW in males (Table 2). The overall model did not predict MCW in males (F=2.2, p=0.059).

In females, all predictors except periodontal status showed significant associations with MCI in the binary logistic regression analysis (Table 3), while multiple logistic regression analysis suggests that T-score, age, and the number of remaining teeth were significantly associated with C2 or C3. A 0.1 SD reduction in T-score increased the odds of having C2 or C3 by 38% (OR 0.62, 95% CI, 0.46; 0.84). Being a decade older resulted in 1.65 (95% CI, 1.11; 2.44) times higher odds of having mildly or severely eroded cortex in females. Every remaining tooth was associated with a significant reduction of 7% in the odds of having C2 or C3 (OR=0.93, 95% CI, 0.87; 0.98). T-score was the strongest predictor of MCI in females, as demonstrated by the Wald statistic value of 9.65. The overall model significantly predicted mildly or severely eroded cortices (p <0.001) and explained 16.3–23.0% of the variation in MCI, as suggested by Cox&Snell and Negelkerke tests.

Binary logistic regression analyses in males show a similar result to females in which all predictors except periodontal status were significantly associated with MCI (Table 3). Multiple logistic regression analysis suggests that T-score remained significantly associated with MCI with an odds ratio of 0.73 (95% CI, 0.56; 0.96). The overall model significantly predicted mildly (C2) or severely (C3) eroded cortex in males (p=0.017) but explained only 4.3-5.8% of the variation in MCI.

Discussion

This study analyzed the relationship among general (sex, age, T-score, menopausal status), local factors (remaining teeth and periodontal status), and morphology of the mandibular cortex assessed by MCW or MCI. All factors together explained the variation in MCW and MCI in females more than in males.

T-score was the strongest predictor of MCW and MCI in females, while in males, T-score was the only significant factor associated with MCI. Previous studies mainly assessed the diagnostic efficacy of radiomorphometric indices for osteoporosis screening [9, 11]. Fewer studies explored the association between MCW, MCI, and BMD in females or males [13, 33, 34]. Analyzing the relationships between radiomorphometric indices and BMD and considering relevant confounding factors is necessary to support using these indices for osteoporosis screening, specifically in males, in whom such studies are lacking. Even though we cannot directly compare the odds ratios and β coefficients from other studies with ours due to different methods and regression models used, it is evident from previous research that females with lower skeletal BMD have thinner and more eroded cortexes [34–37]. A positive correlation between T-score and MCW in females ranging from 0.33 to 0.45 in previous studies was consistent with our findings [34–36]. Our study found no significant association between T-score adjusted for other factors and MCW in males. In

Table 3 The results of binary (unadjusted OR) and multiple (adjusted OR) linear regression analysis predicting MCI in females and males

Predictors	Unadjusted OR, 95% CI	SE	Adjusted OR, 95% CI	SE	Wald statistics	Characteristics of the model
Females						
T-score	0.5 (0.40; 0.64) **	0.12	0.62 (0.46; 0.84) **	0.15	9.65 **	Chi-square = 53.87 p<0.001 Cox&Snell R^2 = 0.163 Negelkerke R^2 = 0.230
Age (10 years increment)	2.61 (2.00; 3.42) **	0.14	1.65 (1.11; 2.44) *	0.20	6.17 *	
Menopausal status	18.17 (5.9; 62.4) **	0.63	3.22 (0.76; 13.54)	0.73	2.55	
Remaining teeth	0.90 (0.87; 0.94) **	0.02	0.93 (0.87; 0.98) *	0.03	5.83 *	
Periodontal status	0.88 (0.62; 1.26)	0.17	0.76 (0.51; 1.13)	0.20	1.78	
Males						
T-score	0.71 (0.56; 0.90) **	0.12	0.73 (0.56; 0.96) *	0.14	5.16 *	Chi-square = 12.01 P = 0.017 Cox&Snell R^2 = 0.043 Negelkerke R^2 = 0.058
Age (10 years increment)	1.36 (1.06; 1.76) *	0.13	1.20 (0.8; 1.65)	0.16	1.26	
Remaining teeth	0.95 (0.92; 0.97) **	0.01	0.97 (0.93; 1.02)	0.02	1.32	
Periodontal status	1.09 (0.77; 1.56)	0.18	0.98 (0.67; 1.43)	0.93	0.00	

*Results are significant at 0.05 level

**Results are significant at 0.001 level

contrast, two previous studies found an association of thin MCW in males with osteoporosis (*T*-score ≤ −2.5 SD) [13, 38]. Unlike our study, Leite et al. also found a correlation coefficient of 0.29 for MCW and BMD at the femoral neck in males [13]. Nevertheless, MCI was associated with *T*-score in males in this study, which was consistent with Leite et al. [13], while another study did not find such an association [38]. This divergence in findings might be related to smaller study samples, differences in sampling procedures, or a lack of adjustment for other factors in previous studies.

Age was another major contributor to MCW and MCI in females but not males. Our results align with previous studies that found an interaction between age and sex in the way that cortical thickness reduced more prominently in females than males [15, 19, 20, 39]. Similar trends were observed in other studies exploring the bone geometry and cortical thickness at different skeletal sites in relation to age and sex [40, 41]. A plausible biological explanation for maintaining mandibular cortical thickness in older men could be sex hormones, which play a crucial role in bone formation and resorption. Estrogens increase endosteal and reduce periosteal bone formation in females during puberty, while androgens accelerate periosteal bone formation in growing males [4, 5]. These physiological mechanisms contribute to sex dimorphism in the adult skeleton. After a certain age, bone resorption exceeds bone formation at the inner bone surface in both sexes. However, bone formation continues at the outer bone surface faster in males than in females due to androgens. Thus, males maintain their cortical bone not because they lose less endosteal bone than women but due to a greater periosteal formation [5]. Our study also found that age was a significant predictor for MCI in females but not males (Tables 2 and 3). On the contrary, several studies found that MCI became more eroded with age, regardless of gender [39, 42, 43]. At the same time, those studies did not consider other confounding factors, which may partly explain the disagreement between ours and previous findings. However, one of the studies was longitudinal, which strengthened their results compared to ours [43].

In our study, the number of remaining teeth was significantly associated with MCW and MCI in females but not males (Tables 2 and 3). It is well established in previous research that mechanical strains and subsequent osteocyte response define the geometry and morphology of skeletal bones [44–46]. Thus, it would be appropriate to hypothesize that the lack of loading forces in edentulous people or those with fewer teeth can independently influence the mandibular cortex. Several studies have found an association between remaining teeth and mandibular cortical morphology, even when controlling for age [16, 19, 20, 42]. Okabe et al. found that the number of remaining teeth was weakly correlated with MCW equally for both sexes (0.19 male and 0.14 female) [14]. Dutra et al. found that the number of remaining teeth was related to the thickness of the mandibular cortex, irrespective of gender [20]. Unlike Okabe et al. and Dutra et al., we did not find relationships between remaining teeth and radiomorphometric indices in males. Our results were consistent with Taguchi et al., who also found no relationships between remaining teeth on both the upper and lower jaws and MCW in males [19]. However, considering only mandibular teeth might be more logical, like some previous studies, since we assess the mandibular cortex [16, 20]. Despite the significant association in females, remaining teeth contributed to the thin and eroded cortex to a minor extent. Similar to our study, Legerton et al. and Gulsahi et al. showed that the influence of dentition on cortical erosion and cortical thickness in the antegonial region was weaker than that of age [16, 42].

In our study, menopausal status was significantly associated with MCW and MCI in univariate analysis but not after adjustments, meaning that the other factors confounded this association greatly. Unlike our results, two previous studies found an association between menopausal status and MCW and MCI but did not consider other relevant confounding factors [47, 48].

The association between periodontitis and mandibular cortical morphology is poorly studied. Recent systematic reviews have shown that many studies have explored the effect of osteoporosis on periodontal health [22, 49]. However, some researchers hypothesized the opposite relationship. Two longitudinal studies explored the independent effect of periodontal disease on skeletal bone tissue and found an increased risk of osteoporosis among people diagnosed with periodontitis after accounting for confounders [50, 51]. The rationale behind those hypotheses was that patients with periodontitis exhibit higher systemic levels of inflammatory mediators such as interleukin (IL-2, IL-6) and tumor necrosis factor (TNF-α) [24]. Those inflammatory mediators affect the remodeling of bone tissue and may thus increase the risk of osteoporosis development [52, 53]. One can also speculate that periodontitis may influence the mandibular cortex via systemic inflammation mechanisms. Tooth loss and mobility due to severe periodontitis leading to a lack of mechanical loading may also influence mandibular cortex morphology.

Unlike few existing studies on the subject, this study found no relationship between mandibular cortical morphology (MCW and MCI) and periodontal status for females or males [47, 48]. That might be due to the difference in study design, the wide variation of periodontal measurements, and case definitions in the literature, which makes comparisons between studies difficult. A recent study exploring the utility of the new periodontal disease classification found that different classification systems affect association estimates in epidemiological studies to a great extent, and the utility of the new classification is not well-studied [54]. The 2017

World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions came to a consensus that staging in periodontitis diagnosis should be based on clinical attachment loss (CAL), RBL, PPD, and other factors like teeth missing due to periodontitis [28]. First, CAL and the reason for tooth loss were unavailable in our study. Had we obtained this additional information, the classification of study participants into the periodontal disease categories would probably have changed. Second, PPD is a parameter that reflects the extent of current inflammatory processes, while RBL reflects patients' periodontal disease experience in the past. It is unclear which parameter is most appropriate when exploring the link between periodontal disease and mandibular cortical morphology. Common sense suggests that cortical bone loss does not occur quickly under an inflammatory process; therefore, RBL as a sign of periodontitis history rather than PPD should be used to explore the abovementioned link. In support of this statement, the systematic review of the association between periodontitis and osteoporosis found that most studies using radiological criteria to define a periodontal case showed significant associations with osteoporosis. At the same time, results were more controversial for studies using clinical measurements for case definition [22].

Another factor that can potentially influence mandibular cortical morphology but has not been included in this study is the mechanical bone load produced by masticatory muscles. It has been previously reported that masticatory load affects both the trabecular and cortical bone in different mandibular regions but predominantly in the angle of the mandible [55–57]. Different types of face anatomy are associated with various masticatory loads: individuals with a short face type (hypo-divergent) have a small mandibular angle and short masticatory muscles with increased masticatory function. In contrast, those with long face types (hyper-divergent) have a large mandibular angle, longer muscles, and decreased masticatory function [58]. A study by Gonca et al., published in 2023, found that individuals with hyper-divergent type exhibited less dense trabecular bone at ramus, condyle, sigmoid notch, and mandibular angle, while mandibular cortical width was thinner only in the second molar projection [59]. Since only DPRs were available in this study, while lateral cephalometric radiographs are commonly used to assess vertical facial type, we could not assess the potential association of masticatory load with the morphology of the mandibular cortex. It might be worthwhile to study this association in future research.

The current study has several limitations. First, it is a cross-sectional study. Despite the arguments regarding the possible links between sex, bone mineral density, age, remaining teeth, periodontitis, and mandibular cortex, we cannot infer the directions of these relationships. Another limitation is the suboptimal inter- and inter-observer agreement of radiomorphometric indices reported in our previous work [32]. Moreover, the panoramic radiographs in this study were not standardized using reference objects, i.e., we could not make precise corrections of MCW for magnification. In addition, the classification of periodontitis is likely to be biased because RBL measurements were performed on DPRs, which distort spatial relationships between anatomical structures to some extent. At the same time, the superimposed cervical spine often hinders bone measurements in the anterior region of the jaws. Suboptimal observer agreement, spatial distortion inherent to DPRs, and image distortion due to the patient's head misalignment were likely to produce random errors and a substantial unexplained variation in mandibular cortical morphology.

Conclusion

T-score followed by age contributed most to variation in MCW and MCI in females, supporting the idea of using MCW and MCI for osteoporosis screening for females. Nonetheless, neither general nor local predictors explained the variation in MCW in males. Only the *T*-score was associated with male MCI, though the association was weaker than in females.

Acknowledgements We are very grateful to Gro Eirin Holde, Ph.D., associate professor at the Department of Clinical Dentistry at UiT, The Arctic University of Norway, and The Public Dental Health Service Competence Centre of Northern Norway (TkNN), for classifying the data on periodontal disease and making them available. We are also grateful to Professor Hans-Peter Műller for his invaluable advice.

Author contribution Conceptualization: Anna Teterina (A.T.), Napat Bolstad (N.B.), Bente Morseth (B.M.). Methodology: A.T., B.M., Sanyalak Niratisairak (S.N.). Validation: A.T., N.B.. Formal analysis: A.T.. Writing an original draft: A.T.. Review: editing and supervision: S.N., B.M., N.B.. Project administration and funding acquisition: N.B. All authors have read and agreed to publish the manuscript. Author contribution. Conceptualization: Anna Teterina (A.T.), Napat Bolstad (N.B.), Bente Morseth (B.M.). Methodology: A.T., B.M., Sanyalak Niratisairak (S.N.). Validation: A.T., N.B.. Formal analysis: A.T.. Writing an original draft: A.T.. Review, editing, and supervision: S.N., B.M., N.B.. Project administration and funding acquisition: N.B. All authors have read and agreed to publish the manuscript.

Funding Open access funding provided by UiT The Arctic University of Norway (incl University Hospital of North Norway) The Arctic University of Norway (UiT), Northern Norway Regional Health Authority (Helse Nord RHF), the University Hospital of North Norway (UNN), and different research funds financed the Tromsø Study. The current study was entirely financed by the Department of Clinical Dentistry, the Faculty of Health Science (UiT).

Data Availability The data supporting this study's findings is available from the Tromsø Study but is not publicly accessible due to licensing restrictions. Researchers affiliated with the institutions with research expertise can access the data upon application to the Data

and Publication Committee for the Tromsø Study (https://uit.no/research/tromsostudy).

Declarations

Ethics approval and consent to participate The Tromsø Study was conducted in accordance with the World Medical Association Declaration of Helsinki [60]. The Regional Committee on Research and Ethics North, REK North (reference 2014/940), and the Norwegian Data Protection Authority (reference 14/01463-4/CGN) approved Tromsø7. In addition, the current study was approved by REK North (reference number 68128) and Norwegian Centre for Research Data (NSD). All the participants gave written informed consent before the Tromsø7 data were collected. The data on each study participant were anonymized and de-identified. All individuals have the right to withdraw their personal data from the Tromsø7.

Conflict of interest The authors declare no competing interests.

References

1. Kanis JA, Cooper C, Rizzoli R, Reginster J-Y (2019) European guidance for the diagnosis and management of osteoporosis in postmenopausal women. Osteoporos Int 30(1):3–44. https://doi.org/10.1007/s00198-018-4704-5
2. Kanis JA, Norton N, Harvey NC, Jacobson T, Johansson H, Lorentzon M et al (2021) SCOPE 2021: a new scorecard for osteoporosis in Europe. Arch Osteoporos 16(1):82. https://doi.org/10.1007/s11657-020-00871-9
3. Epidemiology of osteoporosis and fragility fractures 9, rue Juste-Olivier, CH-1260 Nyon, Switzerland: International Osteoporosis Foundation; [cited 2023 Feb 15]. Available from: https://www.osteoporosis.foundation/facts-statistics/epidemiology-of-osteoporosis-and-fragility-fractures.
4. Callewaert F, Sinnesael M, Gielen E, Boonen S, Vanderschueren D (2010) Skeletal sexual dimorphism: relative contribution of sex steroids, growth hormone-insulin-like growth factor-I (GH-IGF-I) and mechanical loading. J Endocrinol 207(2):127–134. https://doi.org/10.1677/JOE-10-0209
5. Seeman E (2001) Sexual Dimorphism in Skeletal Size, Density, and Strength. J Clin Endocrinol Metabol 86(10):4576–4584. https://doi.org/10.1210/jcem.86.10.7960
6. Weaver CM, Gordon CM, Janz KF, Kalkwarf HJ, Lappe JM, Lewis R et al (2016) The National Osteoporosis Foundation's position statement on peak bone mass development and lifestyle factors: a systematic review and implementation recommendations. Osteoporos Int 27(4):1281–1386. https://doi.org/10.1007/s00198-015-3440-3
7. Klemetti E, Vainio P, Lassila V, Alhava E (1993) Cortical bone mineral density in the mandible and osteoporosis status in postmenopausal women. Eur J Oral Sci 101(4):219–223. https://doi.org/10.1111/j.1600-0722.1993.tb01108.x
8. Morita IDDSP, Nakagaki HDDSP, Taguchi ADDSP, Kato KDDSP, Murakami TDDSP, Tsuboi SDDSP et al (2009) Relationships between mandibular cortical bone measures and biochemical markers of bone turnover in elderly Japanese men and women. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 108(5):777–783. https://doi.org/10.1016/j.tripleo.2009.03.043
9. Calciolari E, Donos N, Park JC, Petrie A, Mardas N (2015) Panoramic measures for oral bone mass in detecting osteoporosis: a systematic review and meta-analysis. J Dent Res 94:17–27. https://doi.org/10.1177/0022034514554949
10. Taguchi A, Tanaka R, Kakimoto N, Morimoto Y, Arai Y, Hayashi T et al (2021) Clinical guidelines for the application of panoramic radiographs in screening for osteoporosis. Oral Radiol 37(2):189–208. https://doi.org/10.1007/s11282-021-00518-6
11. Kinalski MA, Boscato N, Damian MF (2020) The accuracy of panoramic radiography as a screening of bone mineral density in women: a systematic review. Dentomaxillofacial Radiol 49(2):20190149. https://doi.org/10.1259/dmfr.20190149
12. Yeung AWK, Mozos I (2020) The innovative and sustainable use of dental panoramic radiographs for the detection of osteoporosis. Int J Environ Res Public Health 17(7):2449. https://doi.org/10.3390/ijerph17072449
13. Leite AF, de Souza Figueiredo PT, Barra FR, de Melo NS, de Paula AP (2011) Relationships between mandibular cortical indexes, bone mineral density, and osteoporotic fractures in Brazilian men over 60 years old. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 112(5):648–656. https://doi.org/10.1016/j.tripleo.2011.06.014
14. Okabe S, Morimoto Y, Ansai T, Yoshioka I, Tanaka T, Taguchi A et al (2008) Assessment of the relationship between the mandibular cortex on panoramic radiographs and the risk of bone fracture and vascular disease in 80-year-olds. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 106(3):433–442. https://doi.org/10.1016/j.tripleo.2007.09.013
15. Roberts M, Yuan J, Graham J, Jacobs R, Devlin H Changes in mandibular cortical width measurements with age in men and women. Osteoporosis Int. [Internet]. 2011 [cited 2022 Jan 9] 22(6):1915–1925. https://doi.org/10.1007/s00198-010-1410-3
16. Ledgerton D, Horner K, Devlin H, Worthington H Radiomorphometric indices of the mandible in a British female population. Dento-Maxillo-Facial Radiol 28(3):173–181 Available from: https://www.ncbi.nlm.nih.gov/pubmed/107404731999 [cited 2021 Nov 25]
17. Jonasson G, Skoglund I, Rythén M (2018) The rise and fall of the alveolar process: Dependency of teeth and metabolic aspects. Arch Oral Biol 96:195–200. https://doi.org/10.1016/j.archoralbio.2018.09.016
18. Kiliaridis S, Bresin A, Holm J, Strid K-G (1996) Effects of masticatory muscle function on bone mass in the mandible of the growing rat. Cells Tissues Organs 155(3):200–205. https://doi.org/10.1159/000147805
19. Taguchi A, Tanimoto K, Suei Y, Wada T (1995) Tooth loss and mandibular osteopenia. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 79(1):127–132. https://doi.org/10.1016/S1079-2104(05)80088-5
20. Dutra V, Yang J, Devlin H, Susin C (2005) Radiomorphometric indices and their relation to gender, age, and dental status. Oral

Surg Oral Med Oral Pathol Oral Radiol Endod 99(4):479–484. https://doi.org/10.1016/j.tripleo.2004.09.013
21. Hienz SA, Paliwal S, Ivanovski S (2015) Mechanisms of bone resorption in periodontitis. J Immunol Res 2015:615486. https://doi.org/10.1155/2015/615486
22. Martínez-Maestre MÁ, González-Cejudo C, Machuca G, Torrejón R, Castelo-Branco C (2010) Periodontitis and osteoporosis: a systematic review. Climacteric 13(6):523–529. https://doi.org/10.3109/13697137.2010.500749
23. Demmer RT, Trinquart L, Zuk A, Fu BC, Blomkvist J, Michalowicz BS et al (2013) The influence of anti-infective periodontal treatment on C-reactive protein: a systematic review and meta-analysis of randomized controlled trials. PLoS One 8(10):e77441. https://doi.org/10.1371/journal.pone.0077441
24. Hajishengallis G, Chavakis T (2021) Local and systemic mechanisms linking periodontal disease and inflammatory comorbidities. Nat Rev Immunol 21(7):426–440. https://doi.org/10.1038/s41577-020-00488-6
25. Souza PP, Lerner UH (2013) The role of cytokines in inflammatory bone loss. Immunol Investig 42(7):555–622. https://doi.org/10.3109/08820139.2013.822766
26. Hopstock LA, Grimsgaard S, Johansen H, Kanstad K, Wilsgaard T, Eggen AE (2022) The seventh survey of the Tromsø study (Tromsø7) 2015–2016: study design, data collection, attendance, and prevalence of risk factors and disease in a multipurpose population-based health survey. Scand J Public Health 50(7):919–929. https://doi.org/10.1177/14034948221092294
27. Holde GE, Oscarson N, Trovik TA, Tillberg A, Jönsson B (2017) Periodontitis prevalence and severity in adults: a cross-sectional study in norwegian circumpolar communities. J Periodontol 88(10):1012–1022. https://doi.org/10.1902/jop.2017.170164
28. Tonetti MS, Greenwell H, Kornman KS (2018) Staging and grading of periodontitis: Framework and proposal of a new classification and case definition. J Periodontol 89:S159–SS72. https://doi.org/10.1002/JPER.18-0006
29. Trombelli L, Farina R, Silva CO, Tatakis DN (2018) Plaque-induced gingivitis: Case definition and diagnostic considerations. J Clin Periodontol 45:S44–S67. https://doi.org/10.1111/jcpe.12939
30. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9(7):671–675. https://doi.org/10.1038/nmeth.2089
31. Ledgerton D, Horner K, Devlin H, Worthington H Panoramic mandibular index as a radiomorphometric tool: an assessment of precision. Dento-Maxillo-Facial Radiol 26(2):95–100. Available from: [Internet]. 1997 [cited 2021 Nov 25]. https://doi.org/10.1038/sj.dmfr.4600215
32. Teterina A, Niratisairak S, Morseth B, Bolstad N (2023) Diagnostic efficacy of radiomorphometric indices for predicting osteoporosis in a Norwegian population in the Tromsø Study: Tromsø7. Oral Surg Oral Med Oral Pathol Oral Radiol 135(3):444–455. https://doi.org/10.1016/j.oooo.2022.10.039
33. Vlasiadis KZ, Damilakis J, Velegrakis GA, Skouteris CA, Fragouli I, Goumenou A et al (2008) Relationship between BMD, dental panoramic radiographic findings and biochemical markers of bone turnover in diagnosis of osteoporosis. Maturitas. 59(3):226–233. https://doi.org/10.1016/j.maturitas.2008.01.006
34. Leite AF, Figueiredo PT, Guia CM, Melo NS, de Paula AP (2010) Correlations between seven panoramic radiomorphometric indices and bone mineral density in postmenopausal women. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 109(3):449–456. https://doi.org/10.1016/j.tripleo.2009.02.028
35. Marandi S, Bagherpour A, Imanimoghaddam M, Hatef M, Haghighi A (2010) Panoramic-based mandibular indices and bone mineral density of femoral neck and lumbar vertebrae in women. J of Dentist (Tehran, Iran) 7(2):98–106
36. Devlin H, Horner K (2002) Mandibular radiomorphometric indices in the diagnosis of reduced skeletal bone mineral density. Osteoporos Int 13(5):373–378. https://doi.org/10.1007/s001980200042
37. Taguchi A, Suei Y, Ohtsuka M, Otani K, Tanimoto K, Hollender LG (1999) Relationship between bone mineral density and tooth loss in elderly Japanese women. Dentomaxillofacial Radiol 28(4):219–223. https://doi.org/10.1038/sj/dmfr/4600445
38. Dagistan S, Bilge O (2010) Comparison of antegonial index, mental index, panoramic mandibular index and mandibular cortical index values in the panoramic radiographs of normal males and male patients with osteoporosis. Dentomaxillofacial Radiology 39(5):290–294. https://doi.org/10.1259/dmfr/46589325
39. Kavitha MS, Park SY, Heo MS, Chien SI (2016) Distributional variations in the quantitative cortical and trabecular bone radiographic measurements of mandible, between male and female populations of Korea, and its utilization. PLoS One 11(12):e0167992. https://doi.org/10.1371/journal.pone.0167992
40. Wang L, Cheng XG, Su YB, Brown K, Xu L, Li K et al (2017) Sex-related variations in cortical and trabecular bone of the femoral neck in an elderly Chinese population. Osteoporos Int 28(8):2391–2399. https://doi.org/10.1007/s00198-017-4043-y
41. Yates LB, Karasik D, Beck TJ, Cupples LA, Kiel DP (2007) Hip structural geometry in old and old-old age: similarities and differences between men and women. Bone 41(4):722–732. https://doi.org/10.1016/j.bone.2007.06.001
42. Gulsahi A, Yüzügüllü B, Imirzalioglu P, Genç Y (2008) Assessment of panoramic radiomorphometric indices in Turkish patients of different age groups, gender and dental status. Dentomaxillofacial Radiol 37(5):288–292. https://doi.org/10.1259/dmfr/19491030
43. Al-Dam A, Haydar Y, Rashad A (2020) Is there a change in the mandibular cortical shape index in panoramic radiographs over time? A longitudinal retrospective study. Dent Res J (Isfahan) 17(6):465–471
44. Lang T, LeBlanc A, Evans H, Lu Y, Genant H, Yu A (2004) Cortical and trabecular bone mineral loss from the spine and hip in long-duration spaceflight. J Bone Miner Res 19(6):1006–1012. https://doi.org/10.1359/jbmr.040307
45. Heinonen A, Oja P, Kannus P, Sievänen H, Haapasalo H, Mänttäri A et al (1995) Bone mineral density in female athletes representing sports with different loading characteristics of the skeleton. Bone 17(3):197–203. https://doi.org/10.1016/8756-3282(95)00151-3
46. Eastell R, O'Neill TW, Hofbauer LC, Langdahl B, Reid IR, Gold DT et al (2016) Postmenopausal osteoporosis. Nat Rev Dis Primers 2(1):1–16. https://doi.org/10.1038/nrdp.2016.69
47. Mudda JA, Bajaj M, Patil VA (2010) A Radiographic comparison of mandibular bone quality in pre- and post-menopausal women in Indian population. J Ind Soc Periodontol 14(2):121–125. https://doi.org/10.4103/0972-124x.70833
48. Kolte RA, Kolte AP, Potey AM (2017) Risk assessment of osteoporosis in pre- and postmenopausal periodontally healthy and chronic periodontitis women with digital panoramic radiographs. J Indian Soc Periodontol 21(6):461–465. https://doi.org/10.4103/jisp.jisp_238_17
49. Goyal L, Goyal T, Gupta ND (2017) Osteoporosis and periodontitis in postmenopausal women: a systematic review. J Mid-life Health 8(4):151–158. https://doi.org/10.4103/jmh.JMH_55_17
50. Choi JK, Kim YT, Kweon HI, Park EC, Choi SH, Lee JH (2017) Effect of periodontitis on the development of osteoporosis:

results from a nationwide population-based cohort study (2003-2013). BMC Womens Health 17(1):77. https://doi.org/10.1186/s12905-017-0440-9
51. Mau LP, Kuan YC, Tsai YC, Lin JJ, Huynh-Ba G, Weng PW et al (2017) Patients with chronic periodontitis present increased risk for osteoporosis: a population-based cohort study in Taiwan. J Periodontal Res 52(5):922–929. https://doi.org/10.1111/jre.12464
52. Brincat SD, Borg M, Camilleri G, Calleja-Agius J (2014) The role of cytokines in postmenopausal osteoporosis. Minerva Ginecol 66(4):391–407
53. Wei S, Kitaura H, Zhou P, Ross FP, Teitelbaum SL (2005) IL-1 mediates TNF-induced osteoclastogenesis. J Clin Investig 115(2):282–290. https://doi.org/10.1172/jci23394
54. Ke L, Nogueira G, Thomson WM (2023) Influence of case definitions on epidemiological estimates of periodontitis prevalence and its associations with smoking and OHRQoL. Commun Dent Oral Epidemiol 51(2):194–200
55. Koizumi Y, Ishii T, Nishii Y, Nojima K, Sueishi K (2010) Influence of experimental hemi-occlusion on mandibular morphology and internal structure in growing rabbit. Orthodontic Waves 69(2):58–65
56. Guerreiro FS, Diniz P, Carvalho PEG, Ferreira EC, Avancini SRP, Ferreira-Santos RI (2013) Effects of masticatory hypofunction on mandibular morphology, mineral density and basal bone area. Brazil J Oral Sci 12:205–211
57. Masumoto T, Hayashi I, Kawamura A, Tanaka K, Kasai K (2001) Relationships among facial type, buccolingual molar inclination, and cortical bone thickness of the mandible. Eur J Orthodon 23(1):15–23
58. Gomes SGF, Custodio W, Jufer JSM, Cury AADB, Garcia RCMR (2010) Mastication, EMG activity and occlusal contact area in subjects with different facial types, CRANIO®. J Craniomandibul Sleep Pract 28(4):274–279. https://doi.org/10.1179/crn.2010.035
59. Gonca M, Gunacar DN, Kose TE, Beser B (2023) Evaluation of trabecular bone and mandibular cortical thickness in adults with different vertical facial types. Oral Surg Oral Med Oral Pathol Oral Radiol 135(2):282–293
60. World Medical Association (2013) World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. World Med Assoc JAMA 310(20):2191–2194. [cited 2021 Oct 13] 20.Available from:. https://doi.org/10.1001/jama.2013.281053

Paper 3

Paasche Edvardsen, I., Teterina, A., Johansen, T., Myhre, J.N., Godtliebsen, F. & Bolstad, N.L. (2022).

Automatic detection of the mental foramen for estimating mandibular cortical width in dental panoramic radiographs: the seventh survey of the Tromsø Study (Tromsø7) in 2015–2016.

Journal of International Medical Research, 50(11).

Pre-Clinical Research Report

Journal of INTERNATIONAL MEDICAL RESEARCH

Journal of International Medical Research
2022, Vol. 50(11) 1–17

DOI: 10.1177/03000605221135147
journals.sagepub.com/home/imr
SAGE

Automatic detection of the mental foramen for estimating mandibular cortical width in dental panoramic radiographs: the seventh survey of the Tromsø Study (Tromsø7) in 2015–2016

Isak Paasche Edvardsen[1], Anna Teterina[2], Thomas Johansen[3], Jonas Nordhaug Myhre[3], Fred Godtliebsen[4] and Napat Limchaichana Bolstad[2]

Abstract

Objective: To apply deep learning to a data set of dental panoramic radiographs to detect the mental foramen for automatic assessment of the mandibular cortical width.

Methods: Data from the seventh survey of the Tromsø Study (Tromsø7) were used. The data set contained 5197 randomly chosen dental panoramic radiographs. Four pretrained object detectors were tested. We randomly chose 80% of the data for training and 20% for testing. Models were trained using GeForce RTX 2080 Ti with 11 GB GPU memory (NVIDIA Corporation, Santa Clara, CA, USA). Python programming language version 3.7 was used for analysis.

Results: The EfficientDet-D0 model showed the highest average precision of 0.30. When the threshold to regard a prediction as correct (intersection over union) was set to 0.5, the average precision was 0.79. The RetinaNet model achieved the lowest average precision of 0.23, and the

[1]Department of Physics and Technology, Faculty of Science and Technology, UiT The Arctic University of Norway, PO Box 6050, Langnes, N-9037 Tromsø, Norway
[2]Department of Clinical Dentistry, Faculty of Health Sciences, UiT The Arctic University of Norway, PO Box 6050, Langnes, N-9037 Tromsø, Norway
[3]Department of Energy and Technology, NORCE Norwegian Research Center, Sykehusveien 23, 9294 Tromsø, Norway
[4]Department of Mathematics and Statistics, Faculty of Science and Technology, UiT The Arctic University of Norway, PO Box 6050, Langnes, N-9037 Tromsø, Norway

Corresponding author:
Napat Limchaichana Bolstad, Associate Professor of Oral and Maxillofacial Radiology, Department of Clinical Dentistry, Faculty of Health Sciences, UiT The Arctic University of Norway, PO Box 6050, Hansens str., 86, 9019, Langnes, N-9037 Tromsø, Norway.
Email: napat.l.bolstad@uit.no

precision was 0.64 when the intersection over union was set to 0.5. The procedure to estimate mandibular cortical width showed acceptable results. Of 100 random images, the algorithm produced an output 93 times, 20 of which were not visually satisfactory.
Conclusions: EfficientDet-D0 effectively detected the mental foramen. Methods for estimating bone quality are important in radiology and require further development.

Keywords
Dentistry, artificial intelligence, panoramic radiography, machine learning, mental foramen, mandibular cortical width

Date received: 24 June 2022; accepted: 10 October 2022

Introduction

Dental panoramic radiographs (DPRs) are a standard diagnostic tool in dental practice because they provide valuable and comprehensive information about oral health and have a relatively low radiation dose. Approximately 16 million DPRs are annually taken in the general dental service in England and Wales,[1] 10 million in Japan,[2] and 5.55 million in Norway.[3] DPRs provide a comprehensive view of the jaw. In many situations, DPRs assist in providing information on the status of the jaw prior to further examination decisions such as those required in patients with jaw trauma, extensive dental or osseous lesions, tooth eruption, and developmental anomalies.[4]

The mental foramen (MF) is a clinically significant landmark for clinicians in several disciplines, such as dentists, oral and maxillofacial surgeons, emergency physicians, and plastic and reconstructive surgeons.[5] For example, to perform a mental nerve block (a type of anesthesia applied in the region of the MF), accurate determination of the position of the MF is paramount to avoid injury to nerves and blood vessels. The MF is also an essential landmark for measuring the mandibular cortical width (MCW) (Figure 1). A recent systematic review concluded that the MCW measured on DPRs taken for routine dental diagnoses might also be useful as a screening tool for osteoporosis.[6] However, previous studies showed low reliability of the MCW when manually measured by different dentists.[7,8] Therefore, development of an automatic algorithm with which to measure the MCW was proposed.[8] Finding the correct position of the MF is the most important step in building such an automatic algorithm.

The MF is commonly located in the projection of the root apex of the second premolar or between the first and second premolar apices. Irregular tooth alignments or missing teeth make it challenging to determine the location of the MF.[9] Most patients have a single MF. However, variations such as supernumerary (accessory), curling, looping, or missing MFs are also encountered by clinicians. An accessory MF can occur because the mental nerve splits into several nerve fibers before the development of the MF, resulting in double, triple, or quadruple MFs. However, an accessory MF is more common than an absent MF.[9] An accessory MF is present in approximately 1% to 6% of people in different populations. A literature review showed that the MF was

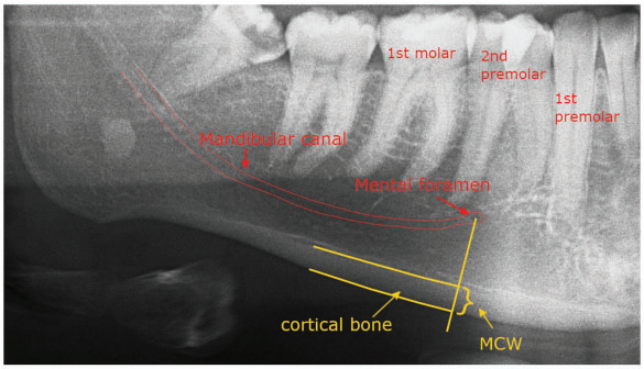


Figure 1. Visualization of a region on a dental panoramic radiograph with essential markings such as the mental foramen, mandibular canal, and cortical bone. The MCW is measured between the border of the bone along the line drawn through the mandibular foramen perpendicular to the tangent of the lower edge of the bone.
MCW, mandibular cortical width.

detectable in approximately 87% to 94% of DPRs but clearly visible in only 49% to 64% of DPRs.[10] Jacobs et al.[11] reported detection of the MF in 94% of 545 DPRs; however, only 49% were considered visible by two independent observers (oral radiologists).

Studies on automatic image analysis from DPRs have been conducted in recent years, and such analysis is challenging because of the inherent complexity of DPRs. The challenge lies in identifying and recognizing specific structures and their morphometry. Morphometry involves assessment of the mandibular cortical bone and MCW for diagnosis of osteoporosis. Before considering an automatic system, Arifin et al.[12] created a manual computer-aided system for measuring the MCW based on gradient analysis of edges in 2006. Because the dentists had to manually determine the position of the MF, Arifin et al.[12] claimed that the experience of the examiners might greatly influence their decision, resulting in poor intra- and inter-examiner agreement. Other studies have focused on automatic segmentation of the mandible.[13–15] The approaches involved techniques such as horizontal integral projections, use of a modified Canny edge detector, morphological operations, thresholding, and use of active contour models. Methods relying on isolation of the cortical bone region are prone to obstacles due to the unclear border of the bone and sometimes its irregular shape. Active contour models, or snakes, require a clear distinction of pixel intensity levels so that the snakes can follow the border of the mandible.[16] Aliaga et al.[17] considered these factors when developing an automatic system for computing mandibular indices in DPRs. The resulting algorithm computed indices inside two regions of interest that tolerated flexibility in sizes and locations, making this process adequately robust. However, they used morphological operations to locate the MF and reported that the proposed approach failed in 5% of 310 cases.[17] Lee et al.[18] used transfer learning for screening osteoporosis in DPRs with a limited data set (680 images). The highest overall accuracy achieved was 84%. Their results showed that transfer learning with pretrained weights and fine-tuning techniques could be helpful and reliable in the automated screening of osteoporosis.

The main objectives of this study were to explore the feasibility of detecting the MF in DPRs with pretrained object detection models and to investigate the possibility of developing an automatic measurement tool of the MCW.

Materials and Methods

Concepts

The main idea behind deep learning is the ability to solve tasks without explicitly designing a rule-based system to do so. Instead, deep learning resolves an assignment by learning from data and adapting to the present task. Hence, the data are often referred to as training data and are essential for proper functioning of deep learning models. A given model that is pretrained and has gained knowledge for a specific task can be further trained to resolve a similar task without the extensive need for data and computing time; e.g., a model trained to recognize apples can be trained to recognize pears. Fine-tuning is an approach of transfer learning that allows implementation of various strategies in which the model is initialized with knowledge (parameters) from a pretrained model. For instance, a model can be initialized with all the parameters from a pretrained model and adjust them regarding the present task, or a selection of these parameters can be set aside and not adjusted for the new task.

Selection of DPRs and image annotation

The data set used in the present study consisted of DPRs taken during the seventh survey of the Tromsø Study (Tromsø7) from 2015 to 2016. The Tromsø Study is a population-based study carried out in repeated cross-sectional surveys.[19] Tromsø7 consisted of a questionnaire-based survey and clinical examinations, including DPRs. The survey enrolled 21,083 participants aged 40 to 99 years.[20] In total, 3951 DPRs were collected following the clinical dental examination (Figure 2). The DPRs were 2821×1376 pixels, were in TIF format, and had 257 dots per inch. Knowing the dots per inch makes it possible to convert between pixels and physical size. In addition, two regions of interest were automatically cropped out for every image at an exact location. The resulting crops were 300×600 (height $\times$ width) pixels. The fixed cropping region did not always capture the jaw because of the varying patient positioning during the examination; such crops were discarded. Distorted images and images with obstructing artifacts were also rejected. Finally, the image was rejected if the experts did not recognize the position of the MF. Of 7902 crops, 5197 were usable (Figure 2), and the MF was annotated by the experts using VIA annotation software.[21] The data were divided into 4157 training images and 1040 test images (Figure 2). Two dentists experienced in oral radiology handpicked 100 "easy images" in which the MF was distinguishable and 101 "complex images" in which the MF was challenging to locate. These handpicked images were used to further analyze the model.

The dentists divided the workload, not annotating the same image to save time. However, to establish the intersection over union (IoU) between them, 706 images were annotated by both experts once. The IoU metric determines the amount of overlap between two boxes compared with their size (Figure 3). True positives are defined based on the IoU being greater than or equal to a threshold (i.e., $IoU\ (\hat{y}^{(i)}, y^{(i)}) > T$, where T is a defined threshold). The IoU between two bounding boxes A and B is defined in Equation 1.

$$\mathrm{IoU}\,(\mathrm{A},\,\mathrm{B}) = \frac{area(A \cap B)}{area(A \cup B)} \qquad (1)$$

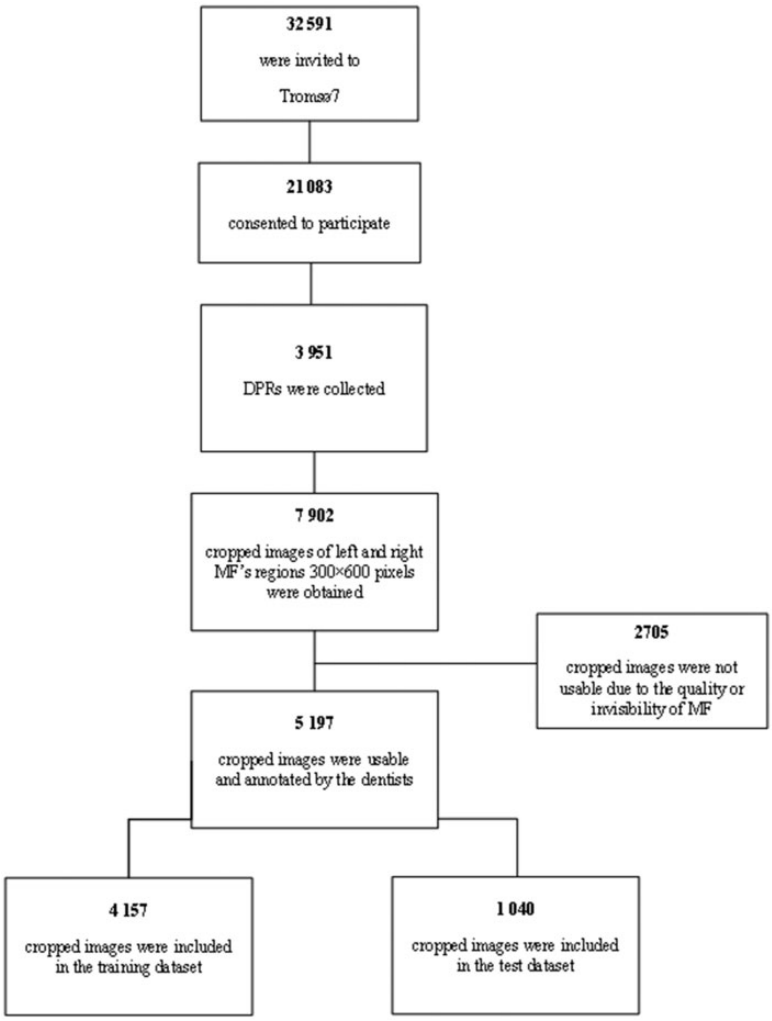


Figure 2. Flow chart of participants included in this study.
DPR, dental panoramic radiograph; MR, mental foramen.

Data availability, ethical permissions, and funding

The current study was based on data owned by the Tromsø Study, Department of Community Medicine, UiT The Arctic University of Norway. The data are available to interested researchers as approved by the Regional Committee for Medical and Health Research Ethics, the Norwegian Data Inspectorate, and the Tromsø Study. Guidelines on data access and the application process are available at https://uit.no/research/tromsostudy.

The Tromsø Study was conducted in accordance with the World Medical Association Declaration of Helsinki.[22] The Regional Committee on Research and Ethics (REK North) and the Norwegian Data Protection Authority (Datatilsynet) approved the Tromsø Study. All participants provided written informed consent.

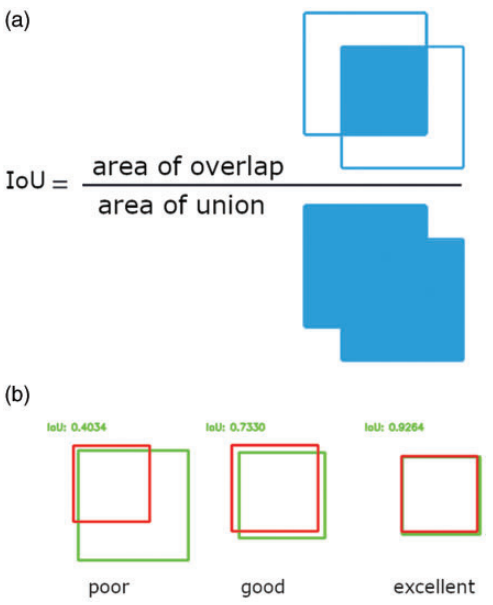


Figure 3. Performance evaluation. (a) Calculation of IoU and (b) poor IoU (0.40), good IoU (0.73), and excellent IoU (0.92). The poor IoU would not be considered a true positive if the threshold was 0.5. IoU, intersection over union.

In addition, we received separate approval from REK North (reference number 68128) and the Norwegian Centre for Research Data (NSD) to use the data from the Tromsø Study database.

The Arctic University of Norway (UiT), Northern Norway Regional Health Authority (Helse Nord RHF), the University Hospital of North Norway (UNN), and different research funds financed the Tromsø Study. The Department of Clinical Dentistry, Faculty of Health Science (UiT) fully financed the current study. We declare no conflict of interest in this study.

Experiment

We performed a feasibility study showing that it is possible to fine-tune an object detector to be adequate in detecting the MF in X-ray images, which is useful for automatic measurement of the MCW. Such a process needs to be able to measure at an appropriate location. Therefore, the first barrier is to detect the MF. The testing and fine-tuning were performed on a GeForce RTX 2080 Ti with 11 GB GPU memory (NVIDIA Corporation, Santa Clara, CA, USA).

The following models, pretrained on the COCO data set,[23] were "fine-tuned" to our data set using the TensorFlow framework:[24]

1. Faster R-CNN with ResNet50[25] as the backbone
2. CenterNet with HourGlass104[26] as the backbone
3. EfficientDet-D0 with EfficientNet-B0[27] as the backbone
4. RetinaNet[28] with ResNet50 as the backbone

Pretrained models (i.e., models that have already been given a data set of input and output pairs and taught to reproduce the correct output for each input) can be useful for solving other tasks involving data that are structured similarly to the original data set. Using pretrained models and training them on a different but similar data set is called fine-tuning. We placed the term "fine-tuning" in quotation marks above because the COCO data set is far from similar to ours, and "trained" hereafter implies "fine-tuned."

Experiment setup

For experiments on object detectors, the I_{0U} threshold ϕ_{IoU} and confidence score threshold ϕ_c used during non-maximum suppression (NMS) were set to 0.5 and virtually 0, respectively, for all models except CenterNet, which does not use NMS. Setting ϕ_c to 0 means all proposals are accepted at the beginning of NMS. We assume that this is beneficial in challenging scenarios in which the predicted scores can be poor. Each model was trained with two

configurations (Setup 1 and Setup 2), and the results are presented in Table 1 and Table 2 (one for each configuration).

The batch size was set to 6 for all experiments (unless something else was specified), and we trained for 30 epochs. Because the training data comprised 4157 examples, processing 6 simultaneously (1 batch) resulted in approximately 693 gradient updates (training steps) to cycle through the training data once (1 epoch). Therefore, training for 30 epochs with a batch size of 6 required approximately 21,000 steps. Empirically, using the moving average of the trained parameters has been shown to be better than using trained parameters directly. However, we did not employ a moving average in any experiment because of technical limitations.

Agreement between different models and dental experts

To evaluate each model, the test images were used to compute the accuracy (i.e., the proportion of images for which the model outputs a correct bounding box). In addition, the handpicked images were used to evaluate the models under the circumstances in which the MF was and was not easy to distinguish. Both experts manually inspected these results because several images were not labeled. The experts reported whether they agreed with the predicted results. The experts performed the inspection of the results once, and the weighted kappa value was calculated.

Procedure to estimate MCW

The procedure to estimate the MCW is briefly described in Algorithm 1. The procedure included the trained object detector. Further, the stop criterion in Algorithm 1 was a user-defined threshold representing the percentage of the line segment L overlapping with black pixels in the binary image I_b (Figure 4). The threshold was set to 0.7 in this study. After Algorithm 1

Table 1. Test results from the object detector with **Experimental Setup 1** of the object detectors presented in the Experiment subsection using the Tromsø7 data set described in the Selection of DPRs and image annotation subsection.

	mAP	mAP at IoU of 0.50	mAP at IoU of 0.75	AR at 100
Faster R-CNN	0.24	0.68	0.069	0.33
CenterNet	0.22	0.68	0.064	0.34
EfficientDet-D0	0.23	0.7	0.007	0.21
RetinaNet	0.21	0.62	0.010	0.46

mAP, mean average precision; IoU, intersection over union; AR, aspect ratio.

Table 2. Test results from the object detector with **Experimental Setup 2** of the object detectors presented in the Experiment subsection using the Tromsø7 data set described in the Selection of DPRs and image annotation subsection.

	mAP	mAP at IoU of 0.50	mAP at IoU of 0.75	AR at 100
Faster R-CNN	0.25	0.72	0.08	0.39
CenterNet	0.28	0.75	0.13	0.39
EfficientDet-D0	0.30	0.79	0.14	0.43
RetinaNet	0.23	0.64	0.01	0.47

mAP, mean average precision; IoU, intersection over union; AR, aspect ratio.

terminated, the width of the bone was defined as the distance between two parallel lines: the initial line and the resulting line. The distance was calculated with Equation 2, where c_1 and c_2 are the y-intercepts of the lines and m is the slope.

Algorithm 1: Method for bone width measurement, which was improved with an object detector. Please see the supplemental material for more information.

Identification of lowest edge of bone

1. Find MF's location P with an object detector
2. Convert image to grayscale and apply median filtering with kernel size 11
3. Apply a variance filter with kernel size 5, and follow with Canny edge detector
4. Use morphology to remove objects smaller than 150 pixels with a neighborhood of 500 pixels
5. Use probabilistic Hough transform[29] to retrieve possible line segments representing the lower bone edge, and save line segment L closest to P

Identification of upper edge of bone (part 1)

1. Convert image to grayscale and apply variance filter with kernel size 8
2. Follow with exposure equalization to obtain I_v
3. Apply a uniform filter with kernel size 11 to I_v to obtain I_m
4. Calculate the binary image

$$I_b = \begin{cases} 1, & \text{if } I_M - I_v \leq \sigma^2 \\ 0, & \text{otherwise} \end{cases}$$

Where $\sigma 2$ is the variance of I_v.

Identification of upper edge of bone (part 2)

- Initialize:

Place line segment L on I_b

while *stop criterion not fulfilled* **do**

| Move L toward P

end

$$d = \frac{|c_1 - c_2|}{\sqrt{1 + \mathrm{m}^2}} \tag{2}$$

Results

The configurations of the hyperparameters of the different algorithms are listed below.

1. Faster R-CNN

Setup 1: The stochastic gradient descent optimizer[30] was used with momentum 0.9 and L_2 regularization $(decay = 4 \times 10^{-4})$. The learning rate grew linearly from 1×10^{-2} to 4×10^{-2} for 2000 steps, then transitioned down using a cosine decay rule.[31] Rectified linear unit activation was employed between convolutional layers. The anchor generator used aspect ratios (1/2, 1, 2) at scales (1/4, 1/2, 1, 2). The training images had a 50% probability of being flipped horizontally.

Setup 2: From the first setup, we changed to the following (the rest was unchanged): Adam optimizer $(\epsilon = 1 \times 10^{-7})$ was used with a learning rate of 2×10^{-4}, which dropped to 2×10^{-4} at epoch 6, then to 8×10^{-5} at epoch 10 and 4×10^{-5} at epoch 15.

2. RetinaNet

***Setup 1*:** The stochastic gradient descent optimizer[30] was used with momentum 0.9 and L_2 regularization $(decay = 4 \times 10^{-4})$.

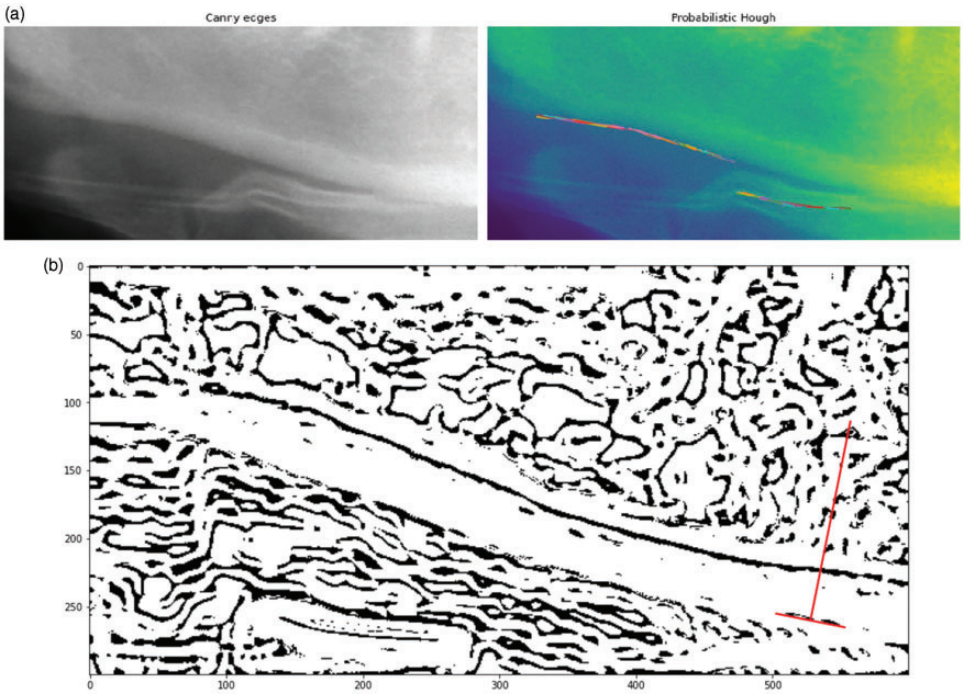


Figure 4. Two cases in which the measuring algorithm needed improvements. (a) Canny edges will be retrieved from the left image and fed to the probabilistic Hough transform to find the best edge candidate. However, an artifact breaks the jawline, and the segment closest to the mental foramen here will be incorrect. (b) A case of a "pit" where the line segment has been initialized on the binary image_1, satisfying the stopping criteria (overlapping black pixels).

The learning rate grew linearly from 1×10^{-2} to 4×10^{-2} for 2000 steps, then transitioned down using a cosine decay rule.[31] Synchronized batch normalization was added after every convolution with batch norm decay of 0.99 with $\epsilon = 1 \times 10^{-3}$. Rectified linear unit activation was employed but was capped at 6. Standard smooth L1 was the localization loss, and focal loss with $\alpha = 0.25$ and $\gamma = 2$ was the classification loss. The anchor generator used aspect ratios (1/2, 1, 2). The training images had a 50% probability of being flipped horizontally. The feature pyramid used minimum level 3 and maximum 7.

Setup 2: From the first setup, we changed to the following (the rest was unchanged): Adam optimizer,[32] where the learning rate grew linearly from 2×10^{-4} to 2×10^{-3} for 2100 steps, then transitioned down using a cosine decay rule.

3. CenterNet

Setup 1: The Adam optimizer was used ($\epsilon = 1 \times 10^{-7}$) for training with a constant learning rate of 9.9×10^{-4}. For the penalty-reduced pixel-wise logistic regression with focal loss, α and β were set to 2 and 4, respectively. The loss was scaled by $\lambda_{size} = 0.1$ and $\lambda_{off} = 1.0$. The training images had a 50% probability of being flipped horizontally, cropped, contrast-adjusted, or brightness-adjusted.

Setup 2: From the first setup, we changed to the following (the rest was

unchanged): The Adam optimizer was used ($\epsilon = 1 \times 10^{-7}$) for training with a learning rate of 5×10^{-4} for 30 epochs, dropping 10× at epochs 18 and 24.

4. EfficientDet-D0

Setup 1: The Adam optimizer ($\epsilon = 1 \times 10^{-7}$) was used with a learning rate of 2×10^{-2} for 30 epochs, dropping 10× at epochs 18 and 24. Synchronized batch normalization was added after every convolution with batch norm decay of 0.99 and $\epsilon = 1 \times 10^{-3}$. Swish-1[33] (commonly called SiLu) activation was employed. Standard smooth L1 was the localization loss, and focal loss with $\alpha = 0.25$ and $\gamma = 1$ was the classification loss. The anchor generator used aspect ratios (1/2, 1, 2, 4). The training images had a 50% probability of being flipped horizontally. The feature pyramid used minimum level 3 and maximum 7.

Setup 2: From the first setup, we changed to the following (the rest was unchanged): Adam optimizer ($\epsilon = 1 \times 10^{-7}$) was used with a learning rate of 2×10^{-4}, which dropped to 1×10^{-4} at epoch 6, then to 8×10^{-5} at epoch 10 and 4×10^{-5} at epoch 15. Random cropping was added as well, and the batch size was increased to 8.

The different performances of the models with respect to detection of the MF are shown in Tables 1 and 2. EfficientDet-D0 clearly performed better in terms of average precision. This was true for both cases in which an IoU of 0.50 and 0.75 was the threshold for a prediction labeled as true positive. The three other models demonstrated relatively fair results. Notably, EfficientDet-D0 only uses a fraction of the number of parameters compared with the other models. However, CenterNet produced very similar results, and RetinaNet had a higher average recall regarding 100 detections. In addition, we noticed that the second configuration of every model produced better mean average precision than the first.

The initial hypothesis of this study was that existing models trained on the COCO data set could be fine-tuned to detect the MF. EfficientDet-D0 demonstrated sufficient precision and correct predictions at a threshold of 50% IoU compared with the other well-known models tested in this study; thus, the first hypothesis was concluded to be true. This conclusion was drawn by comparing the average precisions in Tables 1 and 2.

Figures 5, 6, 7, and 8 show the agreement between the dental experts when assessing the results of automatic detection of the MF on the handpicked images. This further investigation showed that the experts agreed on every prediction using the easy images and disagreed on some of the more complex images (Tables 3 and 4). It is apparent that annotating complex images is exceptionally challenging, and in the worst cases, annotation relies only on the best guess. When using three categories ("agree," "unsure," and "disagree"), the kappa value was 0.18, indicating slight agreement.[34] However, the kappa value can be misleading when the distribution between categories is unequal,[35] as in our case where only 10 of 101 predictions fell into the category "disagree."

The second hypothesis followed the first, assuming the first was true: Can an object detector help accomplish automatic measurement of the MCW in DPRs? Using the results obtained from testing the first hypothesis, it was possible to make an algorithm that automates the measuring process. Of 100 random images (not necessarily in the training or test data set), the algorithm produced an output 93 times, 20 of which were not visually satisfactory. Therefore, the resulting algorithm needs improvement, and it is not yet generalized to handle image regions with high complexity even though the MF was found.

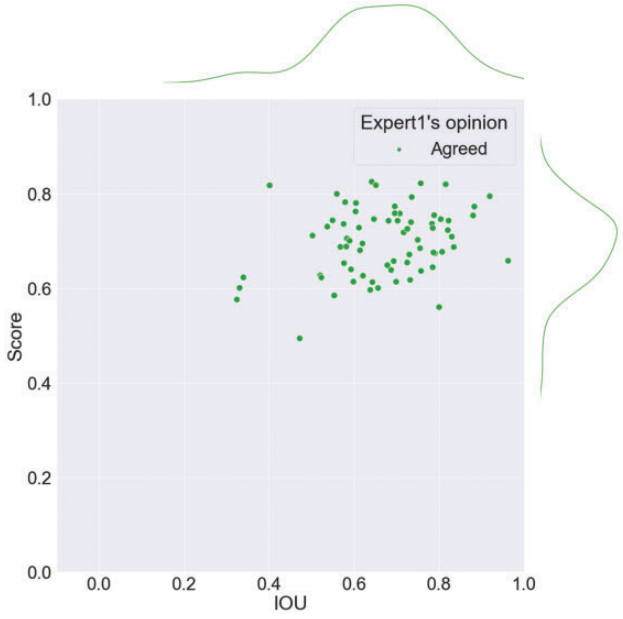


Figure 5. Predicted score versus IoU. Expert 1 has manually inspected the results and indicated whether they agree with the predicted results.
IoU, intersection over union.

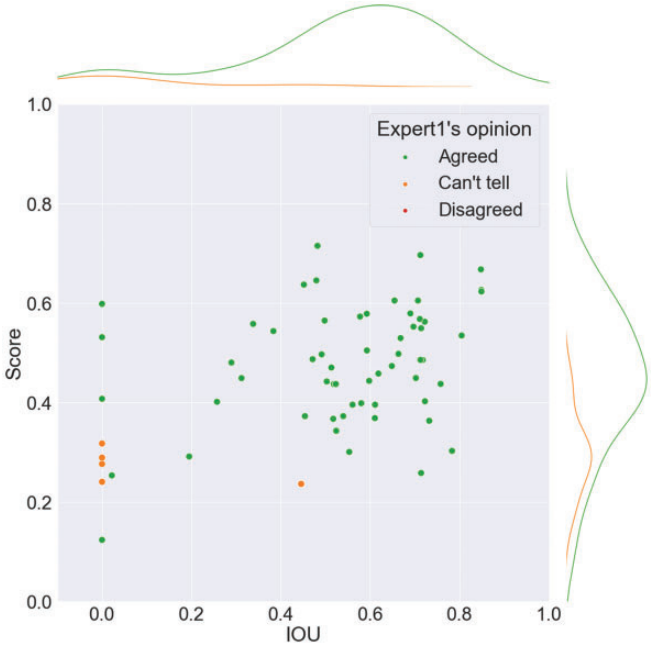


Figure 6. Predicted score versus IoU. Expert 1 has manually inspected the results and indicated whether they agree with the predicted results.
IoU, intersection over union.

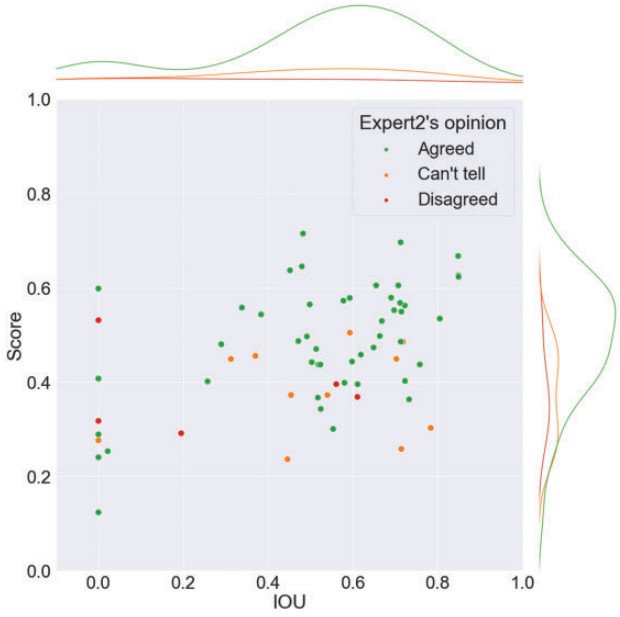


Figure 7. Predicted score versus IoU. Expert 2 has manually inspected the results and indicated whether they agree with the predicted results.
IoU, intersection over union.

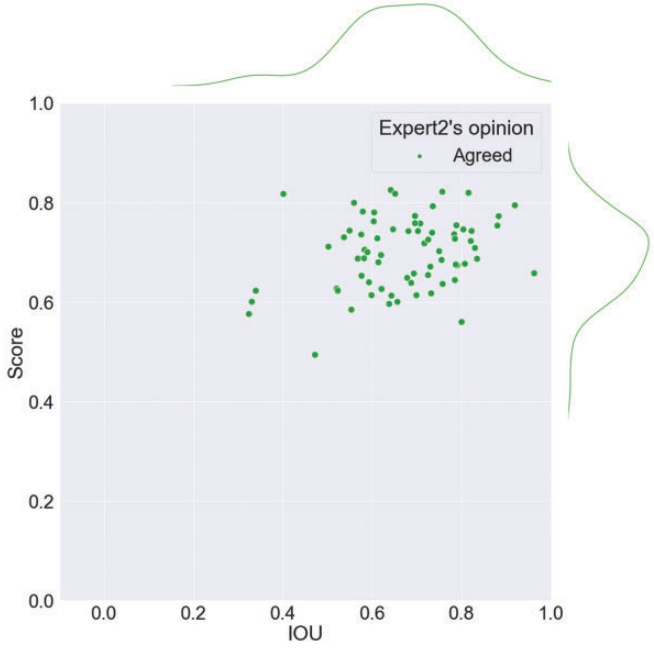


Figure 8. Predicted score versus IoU. Expert 2 has manually inspected the results and indicated whether they agree with the predicted results.
IoU, intersection over union.

Table 3. Evaluation of 101 complex images by two dentists.

	Expert 2 (agree)	Expert 2 (unsure)	Expert 2 (disagree)
Expert 1 (agree)	67	12	7
Expert 1 (unsure)	7	5	2
Expert 1 (disagree)	0	1	0

Table 4. Combined evaluation of 101 complex images.

	Expert 1 (agree)	Expert 1 (disagree)
Expert 2 (agree)	91	7
Expert 2 (disagree)	0	3

Therefore, the algorithm was semi-capable of measuring the bone from visual reports, and the second hypothesis cannot be considered true.

Discussion

Our investigation of the predictive ability of EfficientDet-D0 using easy images showed that both experts agreed with every prediction, even when several predictions had a relatively low IoU (<0.5). However, even when the IoU was poor, overlap was still present between the ground truth and the predicted bounding box. Consequently, the prediction can result in a good suggestion of the position of the MF that largely agrees with dental experts.

Analysis of the prediction ability of EfficientDet-D0 using complex images produced some interesting results. First, it should be stated that ground truths are not absolute. The experts agreed or were unsure about predictions for which the IoU was 0. All of these predicted bounding boxes lay on the mandibular canal next to the ground truth. Therefore, these predictions possibly contained the MF. The expert verdict explained that other predicted regions seemed to contain part of the tooth's root apex, which could be a dark region in some cases and is challenging to distinguish from the MF.

Another case (Figure 7) disagreed with two predictions with a relatively high IoU (>0.5), which may seem contradictory. This shows that the cropped images were challenging to label with a ground truth bounding box; labeling could only be accomplished by the best guess. Additionally, the entire image was available to aid the evaluation of a prediction in cases where the cropped images lacked information on other important landmarks, such as the premolars, which might explain this scenario. If no other landmarks are present when evaluating a prediction of the location of the MF, explainable artificial intelligence (AI) is needed to provide insight into the reason behind the predictions.[36] This would also allow for an uncertainty measure behind the model, which would benefit clinicians.

As stated above, not all the complex images that were handpicked for inference had ground truth bounding boxes. This occurred because the experts could not locate the MF when creating ground truth bounding boxes. These highly complex images were given to the model, and the experts evaluated the results (see Table 3). In one case, one expert disagreed with the prediction whereas the other expert was unsure. In another case, one expert was unsure but leaned toward disagreeing whereas the other expert disagreed with the prediction. These cases are depicted in Figures 9(a) and (b). For all cases shown in Figures 9(a) and (b), the experts concluded that the model annotated a part of the tooth's root apex, or the experts could not see the MF and therefore disagreed.

In this study, EfficientDet-D0 was used for inference, while EfficientDet-D7 is available with almost twice the mean average precision on the COCO data set. Future

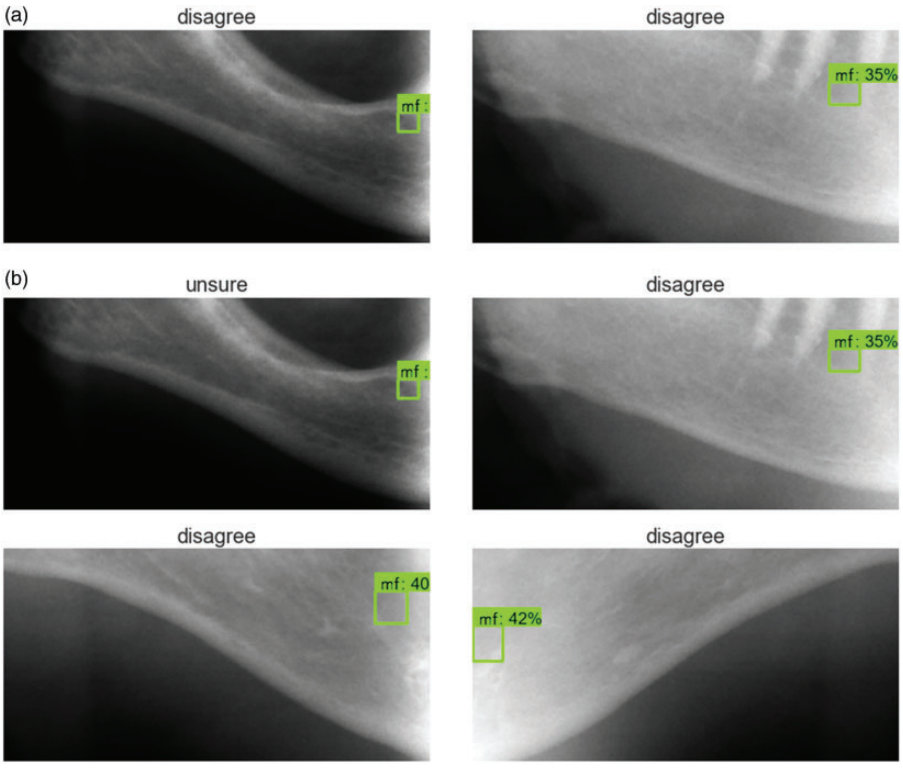


Figure 9. Incorrect prediction from EfficientDet-D0 as judged by the (a) first and (b) second experts.
MF, mental foramen.

research should utilize explainable AI to improve the trustworthiness of the AI system.[36]

When estimating the MCW, the proposed Algorithm 1 operates *fully* automatically given an image region. In Figure 10, we see that the algorithm effectively locates the MF and estimates the bone thickness automatically. However, our study did not compare automated MCW measurements with the actual osteoporosis status based on hip bone mineral density. Unlike our study, the OSTEODENT study used active shape models for automated MCW measurements and compared them with the actual diagnoses. The authors found that an MCW of <3 mm could identify postmenopausal women with osteoporosis and stated that their findings were clinically important.[37] Thus, in our further study, we plan to determine whether the algorithm measuring MCW can differentiate patients with osteoporosis diagnosed by bone mineral density measurements at the hip.

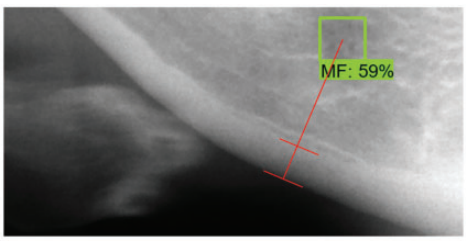


Figure 10. Results from Algorithm 1. The optimistic results are observed in this radiograph. The algorithm has stopped in a sweet spot immediately under porous textures.
MF, mental foramen.

Moreover, to further improve the MCW measurement algorithm, steps can be taken to check whether the cropped image contains the lower edge of the bone beneath the MF. Alternatively, a dynamic image-cropping procedure based on other landmarks could be implemented to ensure the presence of the edge. Otherwise, the algorithm measures other structures close to the MF, not the bone. Another issue to consider is that the initial lines can become stuck in a "pit" in the binary image I_b (see Algorithm 1) if the lower border of the bone is unclear. In the most challenging scenario, the binary image I_b can contain artifacts overlapping either the line's pathway when traveling toward the MF or other image areas. These artifacts cause an unclear upper bone border, terminating the algorithm at an incorrect location, or the line segment suggested in the first place will suffer (see Figure 4(a)). Therefore, we should also consider possibilities other than the MCW for screening osteoporosis, especially transfer learning, which could be used to learn attributes of DPRs labeled as affected, given a sufficiently large data set.

The use of AI in medicine and dentistry aims at smooth integration into the workflow and saving of time. However, one limitation of AI is that its accuracy depends on the quality of data from which the algorithm has learned. If a human decision is used as a "ground truth," common human bias can be introduced into the algorithm. In this study, expert assessments were considered a "ground truth." The proper "ground truth" for the location of the MF should be either a cadaver mandible or a cone-beam computed tomography scan. However, the former would not be approved by an ethics committee, and the latter was unavailable for our study.

Moreover, medical images with multiple overlapping artifacts can lead to unreliable algorithm outputs, which is also a limitation.[38] Very few studies to date have focused on the automated location of the MF on DPRs. Discussing our findings in the context of previous research is challenging because of the different AI methods used[39] and the lack of guidelines for comparing different studies using AI in medicine, notably dentistry.[38]

Conclusion

The MF is an important landmark for dental practitioners. Detecting its location on a DPR is the most important step in measuring the MCW, which can be a useful index for osteoporosis screening. In this study, EfficientDet-D0 showed sufficient precision and correct predictions of MF locations. Moreover, it was possible to merge EfficientDet-D0 with the previously made MCW measurement algorithm. This indicates the feasibility of fully automatic measurement of the MCW for osteoporosis detection.

Authors' contributions

Isak Paasche Edvardsen: Conceptualization, Methodology, Validation, Formal analysis, Writing – Original Draft.
Anna Teterina: Methodology, Writing – Review, Editing, Visualization, Supervision.
Thomas Johansen, Jonas Nordhaug Myhre, Fred Godtliebsen: Conceptualization, Methodology, Validation, Writing – Review, Editing, Supervision.
Napat Limchaichana Bolstad: Conceptualization, Validation, Writing – Review, Editing, Supervision, Project Administration, Funding Acquisition.

Declaration of conflicting interests

The authors declare that there is no conflict of interest.

Funding

The authors disclosed receipt (pending publication) of the following financial support for the

research, authorship, and/or publication of this article: UiT The Arctic University of Norway, Northern Norway Regional Health Authority (Helse Nord RHF), the University Hospital of North Norway (UNN), and different research funds financed the Tromsø Study. The Department of Clinical Dentistry, Faculty of Health Science, UiT The Arctic University of Norway fully financed the current study.

ORCID iDs

Thomas Johansen https://orcid.org/0000-0003-3572-4706
Napat Limchaichana Bolstad https://orcid.org/0000-0002-4276-6720

Supplemental material

Supplemental material for this article is available online.

References

1. Rushton VE and Horner K. The use of panoramic radiology in dental practice. *J Dent* 1996; 24: 185–201. DOI: 10.1016/0300-5712(95)00055-0.
2. Taguchi A, Tsuda M, Ohtsuka M, et al. Use of dental panoramic radiographs in identifying younger postmenopausal women with osteoporosis. *Osteoporos Int* 2006; 17: 387–394. DOI: 10.1007/s00198-005-2029-7.
3. Saxebøl G and Olerud HM. *Radiation use in Norway. Useful use and good radiation protection for society, humans and the environment*. 2014. Statens strålevern.
4. White SC and Pharoah MJ. *Oral radiology: principles and interpretation*. 7th ed. St. Louis: Elsevier, 2014.
5. Laher AE, Wells M, Motara F, et al. Finding the mental foramen. *Surg Radiol Anat* 2016; 38: 469–476. DOI: 10.1007/s00276-015-1565-x.
6. Calciolari E, Donos N, Park JC, et al. Panoramic measures for oral bone mass in detecting osteoporosis: a systematic review and meta-analysis. *J Dent Res* 2015; 94: 17S–27S. DOI: 10.1177/0022034514554949.
7. Ledgerton D, Horner K, Devlin H, et al. Radiomorphometric indices of the mandible in a British female population. *Dentomaxillofac Radiol* 1999; 28: 173–181. DOI: 10.1038/sj.dmfr.4600435.
8. Devlin C, Horner K and Devlin H. Variability in measurement of radiomorphometric indices by general dental practitioners. *Dentomaxillofac Radiol* 2001; 30: 120–125.
9. Hasan T. Mental foramen morphology: a must know in clinical dentistry. *J Pak Dent Assoc* 2012; 21: 167–172.
10. Greenstein G and Tarnow D. The mental foramen and nerve: clinical and anatomical factors related to dental implant placement: a literature review. *J Periodontol* 2006; 77: 1933–1943. DOI: 10.1902/jop.2006.060197.
11. Jacobs R, Mraiwa N, Van Steenberghe D, et al. Appearance of the mandibular incisive canal on panoramic radiographs. *Surg Radiol Anat* 2004; 26: 329–333. DOI: 10.1007/s00276-004-0242-2.
12. Arifin AZ, Asano A, Taguchi A, et al. Computer-aided system for measuring the mandibular cortical width on dental panoramic radiographs in identifying postmenopausal women with low bone mineral density. *Osteoporos Int* 2006; 17: 753–759.
13. Abdi AH, Kasaei S and Mehdizadeh M. Automatic segmentation of mandible in panoramic X-ray. *J Med Imag* 2015; 2: 044003. DOI: 10.1117/1.JMI.2.4.044003.
14. Kavitha MS, Asano A, Taguchi A, et al. The combination of a histogram-based clustering algorithm and support vector machine for the diagnosis of osteoporosis. *Imaging Sci Dent* 2013; 43: 153–161. DOI: 10.5624/isd.2013.43.3.153.
15. Naik A, Vinayak Tikhe S, Bhide SD, et al. Designing a feature vector for statistical texture analysis of mandibular bone. *Indian J Sci Technol* 2016; 9: 1–4. DOI: 10.17485/ijst/2016/v9i33/96305.
16. Cootes TF, Taylor CJ, Cooper DH, et al. Active shape models-their training and application. *Comput Vis Image Underst* 1995; 61: 38–59.
17. Aliaga I, Vera V, Vera M, et al. Automatic computation of mandibular indices in dental panoramic radiographs for early osteoporosis detection. *Artif Intell Med* 2020; 103: 101816. DOI: 10.1016/j.artmed.2020.101816.

18. Lee K-S, Jung S-K, Ryu J-J, et al. Evaluation of transfer learning with deep convolutional neural networks for screening osteoporosis in dental panoramic radiographs. *J Clin Med* 2020; 9: 392. DOI: 10.3390/jcm9020392.
19. Jacobsen BK, Eggen AE, Mathiesen EB, et al. Cohort profile: the Tromsø Study. *Int J Epidemiol* 2012; 41: 961–967. DOI: 10.1093/ije/dyr049.
20. The seventh survey of the Tromsø study (Tromsø7), https://uit.no/research/tromsoundersokelsen/project?p_document_id=705235&pid=706786 (2022, accessed 27 April).
21. Dutta A and Zisserman A. The VIA Annotation Software for images, audio and video. In: *MM '19: Proceedings of the 27th ACM International Conference on Multimedia*, Nice, France, 21–25 October 2019, pp.2276–2279. DOI: 10.1145/3343031.3350535.
22. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA: World Medical Association* 2013; 310: 2191–2194.
23. Lin TY, Maire M, Belongie S, et al. *Microsoft COCO: Common Objects in Context*. Cham: Cham: Springer International Publishing, 2014, pp.740–755.
24. Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-scale machine learning on heterogeneous systems, https://www.tensorflow.org/ (2015, accessed 30 May 2022).
25. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27–30 June 2016, pp.770–778. DOI: 10.1109/CVPR.2016.90.
26. Newell A, Yang K and Deng J. *Stacked hourglass networks for human pose estimation*. Cham: Cham: Springer International Publishing, 2016, pp.483–499.
27. Tan M and Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. In: *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, CA, USA, 9–15 June 2019, pp.6105–6114.
28. Lin TY, Goyal P, Girshick R, et al. Focal loss for dense object detection. In: *Proceedings of the IEEE international conference on computer vision* 2017, pp.2980–2988.
29. Kiryati N, Eldar Y and Bruckstein AM. A probabilistic Hough transform. *Pattern Recognit* 1991; 24: 303–316. DOI: 10.1016/0031-3203(91)90073-E.
30. Qian N. On the momentum term in gradient descent learning algorithms. *Neural Netw* 1999; 12: 145–151. DOI: 10.1016/S0893-6080(98)00116-6.
31. Loshchilov I and Hutter F. SGDR: Stochastic Gradient Descent with Warm Restarts.
32. Kingma DP and Ba J. Adam: A Method for Stochastic Optimization.
33. Ramachandran P, Zoph B and Le QV. Searching for Activation Functions.
34. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174. DOI: 10.2307/2529310.
35. Szklo M and Nieto FJ. *Epidemiology: Beyond the Basics*. 4 ed. Sudbury: Sudbury: Jones & Bartlett Learning, LLC, 2018.
36. Adadi A and Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access* 2018; 6: 52138–52160.
37. Devlin H, Allen P, Graham J, et al. The role of the dental surgeon in detecting osteoporosis: the OSTEODENT study. *Br Dent J* 2008; 204: E16–E561. DOI: 10.1038/sj.bdj.2008.317.
38. Pethani F. Promises and perils of artificial intelligence in dentistry. *Aust Dent J* 2021; 66: 124–135. DOI: 10.1111/adj.12812.
39. Kats L, Vered M, Blumer S, et al. Neural network detection and segmentation of mental foramen in panoramic imaging. *J Clin Pediatr Dent* 2020; 44: 168–173.