Semi-Automated veRtebral fracture Assessment in ChildrEN

SARACEN

By

Fawaz F Alqahtani

Thesis Submitted to the University of Sheffield for the Degree of

Doctor of Philosophy

Department of Oncology and Metabolism

Medical School

April 2019

Abstract

Aim To evaluate the diagnostic accuracy and reliability in children, of morphometric vertebral fracture analysis (MXA) by using semi-automated 6-point (SpineAnalyzer™) and 33-point (AVERT™) software techniques developed for vertebral fracture diagnosis in adults, which record percentage loss of vertebral body height.

Materials and Methods (i) Retrospective analysis of 137 paediatric lateral spine radiographs (T4 to L4) by five observers using SpineAnalyzer™. (ii) Retrospective analysis of 100 (50 dual-energy x-ray absorptiometry (DXA) and 50 radiographic) images collected from Sheffield Children's Hospital, by two observers using SpineAnalyzer™ and AVERT™. For (i) and (ii) a previous consensus read of radiographs by three paediatric radiologists using a simplified algorithm-based qualitative technique served as the reference standard. (iii) 420 lateral spine DXA were retrospectively collated of children aged between 5 and 18 years. Vertebral fracture assessment (VFA) was performed by an expert paediatric radiologist and served as the reference standard. For (i), (ii) and (iii), diagnostic accuracy (sensitivity, specificity, false negative (FN) and positive (FP) rates) was calculated. Inter and intraobserver agreement levels were calculated using the kappa statistic.

Results For (i) and (ii) Low diagnostic accuracy and poor inter and intraobserver agreement were obtained. For (iii) Overall sensitivity, specificity, FP and FN rates using MXA were 89%, 79%, 21%, and 11% respectively, but for mild fractures alone were 36%, 86%, 14%, and 64% respectively. MXA reached only moderate agreement when compared to the visual semi-quantitative VFA technique, with fair to moderate inter and intraobserver agreement.

Conclusions Neither AVERT™ nor SpineAnalyzer™ is satisfactorily reliable for vertebral fracture diagnosis in children. In order to facilitate the detection of mild vertebral fractures in children, a paediatric standard is required which not only incorporates specific vertebral body height ratios but also the age-related physiological changes in vertebral shape that occur throughout childhood.

To my family.

My parents – For their love, caring and prayers

My wife and daughter – For their usual love and constant encouragement

My brothers and sisters – For their support

Table of contents

List of figures

List of tables

List of appendices

Appendix 1 – Abstracts, Posters and Awards

Appendix 2 – Publications

Appendix 3 – Permissions

List of abbreviations

Abbreviation	Definition
ABQ	Algorithm based qualitative
BMC	Bone mineral content
BMD	Bone mineral density
CT	Computed tomography
DICOM	Digital imaging and communications in medicine
DXA	Dual-energy x-ray absorptiometry
FN	False negative
FP	False positive
HCS	Hierarchical clustering-based segmentation
HR- pQCT	High-resolution peripheral quantitative computed tomography
ICC	Intraclass correlation coefficient
ISCD	International Society for Clinical Densitometry
JPEG	Joint Photographic Experts Group
MRI	Magnetic resonance imaging
MXA	Morphometric vertebral fracture analysis
PACS	Picture archiving and communication system
PhD	Doctor of philosophy
pQCT	Peripheral quantitative computed tomography
QCT	Quantitative computed tomography
RFRV-CLMs	Random forest regression voting constrained local models
SARACEN	Semi-Automated vertebral fRacture Assessment in ChildrEN
SCH	Sheffield Children's Hospital
SID	Source to image distance
SQ	Semi-quantitative
SSMs	Statistical shape models
UK	United Kingdom
UoM	University of Manchester
UoS	University of Sheffield
US	Ultrasound
VFA	Vertebral fracture assessment
XR	Radiograph

Declaration

I hereby declare that this thesis has been composed by myself and has not been accepted in any previous application for a higher degree. The work reported in this thesis is novel and has been carried out by me, with all sources of information being specifically acknowledged by means of references and acknowledgements.

Fawaz Alqahtani
April 2019

Conferences and publications

Alqahtani, F.F., Messina, F., Kruger, E., Gill, H., Ellis, M., Lang, I., Broadley, P. and Offiah, A.C., 2017. Evaluation of a semi-automated software programme for the identification of vertebral fractures in children. *Clinical radiology*, *72*(10), pp.904-e11. DOI.org/10.1016/j.crad.2017.04.010

Alqahtani, F.F and Offiah, A.C, 2018. Diagnosis of osteoporotic vertebral fractures in children. *Pediatric radiology*, available Online: DOI: 10.1007/s00247-018-4279-5

Alqahtani, F.F., Messina, F., Kruger, E., Gill, H., Ellis, M., Lang, I., Broadley, P. and Offiah, A.C., Evaluation of the SpineAnalyzer™ software programme on paediatric radiographic images. United Kingdom Radiological Congress (UKRC), June 2016, Liverpool, UK, Poster presentation

Alqahtani, F.F., Messina, F., Kruger, E., Gill, H., Ellis, M., Lang, I., Broadley, P. and Offiah, A.C., Assessment of a semi-automated software programme for the identification of vertebral fractures in children. The 8th International Conference on children's Bone Health (ICCBH), June 2017, Wurzburg, Germany, Poster presentation, *Bone Abstracts* (2017) 6 ,P 079. DOI: 10.1530/boneabs.6.P079

Alqahtani, F.F., Crabtree, N.J., Bromiley P., Cootes T., Broadley, P., Lang, I. and Offiah, A.C., Diagnostic performance of morphometric vertebral fracture analysis (MXA) using a 33-point software programme on DXA images of children. The 9th International Conference on children's Bone Health (ICCBH), June 2019, Salzburg, Austria, Poster presentation.

Alqahtani, F.F., Messina, F. and Offiah, A.C., Evaluation of a 33-point software programme for the identification of vertebral fractures in children. The 8th Annual Mellanby Centre Research Day, December 2017. Sheffield, UK, Poster presentation

Alqahtani, F.F., Messina, F. and Offiah, A.C., Evaluation of a 33-point software programme for the identification of vertebral fractures in children. The 30th European Congress of Radiology (ECR), March 2018. Vienna, Austria. Oral presentation. *Insights into Imaging Abstracts* (2018). DOI.org/10.1007/s13244-018-0603-8

Alqahtani, F.F., Crabtree, N.J., Bromiley P., Cootes T., Broadley, P., Lang, I. and Offiah, A.C., Diagnostic performance of morphometric vertebral fracture analysis (MXA) using a 33-point software programme on DXA images of children. The 9th Annual Mellanby Centre Research Day, November 2018. Sheffield, UK, Poster presentation

Alqahtani, F.F., Crabtree, N.J., Bromiley P., Cootes T., Broadley, P., Lang, I. and Offiah, A.C., Diagnostic performance of morphometric vertebral fracture analysis (MXA) using a 33-point software programme on DXA images of children. Mellanby centre internal seminar, March 2019, Sheffield, UK. Oral presentation

The published articles were reused in this thesis with the kind permission of the copyright holders. (Appendix 3, page 181)

Achievement and Awards

New Investigator Award, at 8th International Conference on Children's Bone Health (ICCBH) 10-13 June 2017, Wurzburg, Germany (Appendix 1, page 173)

Chapter One
Introduction

1. Introduction

1.1 Motivation

Fractures are a common occurrence during childhood. Indeed, some children are born with bone problems or skeletal diseases which make their bones weaker. In these cases, they are more likely to develop fractures, including vertebral fractures. These vertebral fractures are different from osteoporotic fractures of the limbs in the sense that they are normally silent and if they are not treated, progressive degeneration of vertebral body height occurs along with development of scoliosis. However, if diagnosed early, medical care can be initiated (such as providing treatment with bisphosphonates) to limit the progression of fractures.

There are a number of factors that have motivated the research undertaken in this PhD thesis. Over the past 20 years, much research has been conducted on osteoporotic vertebral fracture in adults, but rather less effort has been devoted to paediatric vertebral fractures. Ultimately, this study aims to increase our understanding of this demographic. Additionally, there is no accepted standardised technique for detecting vertebral fractures in children as yet, which depends to a large degree on identifying loss of vertebral body height and modifications in shape. Traditionally, osteoporotic vertebral fractures are diagnosed from lateral spine radiographs; however, a small number of studies have shown that dual energy x-ray absorptiometry (DXA) is comparable to radiographs for identifying vertebral fractures in children, allowing reduced radiation exposure. Vertebral fracture assessment (VFA) is the term given to diagnose vertebral fractures using DXA. It should be noted that the diagnosis of vertebral fractures in children is highly dependent on the experience of the radiologist reviewing the lateral radiograph and/or VFA, and whilst it is relatively easy to diagnose severe vertebral fracture, mild and moderate fractures are harder to recognise. The

reasons behind this difficulty include the lack of an objective reference standard for normal vertebral shape, and the absence of a standardised scoring system. This latter reason has caused significant variability of diagnosis between different observers when diagnosing vertebral fractures in children. To address these issues, an objective tool that (when compared to the subjective and semi-quantitative methods presently available) can allow more valid and accurate identification of vertebral fractures in children is required. Semi-automated software programmes, such as SpineAnalyzer™ (Optasia Medical, Cheadle, UK) and AVERT™ (Optasia Medical, Cheadle, UK) may be the solution, but so far, limited studies have been carried out to evaluate morphometric vertebral fracture analysis (MXA) using these programmes in children. Importantly, such tools will allow non-radiologist readers, who may not have advanced training in radiology (technical staff and/or physicians), to perform the semi-automated vertebral morphometry measurements, thereby freeing the radiologist from such work.

1.2 Aims and objectives

The overarching objective of this study is to enhance the assessment of vertebral fractures in children by determining the clinical use of semi-automated software programmes that already exist for adults in terms of their success in the identification of vertebral fractures in children.

To achieve this overall aim, the sub-aims and objectives of this PhD are as follows:

1. To describe osteoporosis in children, highlight current diagnostic techniques for vertebral fracture diagnosis, and to outline the different scoring systems available for recording severity of vertebral fractures in children.

2. To assess the accuracy and reliability, when used on children, of the semi-automated 6-point technique (SpineAnalyzer™ software, version number:

4.0.2.19) developed for vertebral fracture diagnosis in adults, which records percentage loss of vertebral body height.

3. To assess whether observer reliability and diagnostic accuracy of MXA for the identification of vertebral fracture in children are improved by using a 33-point semi-automated programme (a novel programme called AVERT™) compared to the 6-point programme (SpineAnalyzer™).

4. To evaluate MXA using AVERT™ on a large cohort of children with chronic disease. More specifically, using the latest iDXA imaging technology and comparing with using the visual semi-quantitative (SQ) method by an experienced paediatric radiologist for the identification of vertebral fractures in children.

5. To assess inter and intraobserver agreement of MXA independently performed by three consultant paediatric musculoskeletal radiologists and an experienced clinical scientist using AVERT™.

1.3 Outline of thesis

This thesis has been written and presented as an alternative format thesis (suitable for submission for publication in a peer-reviewed journal). The reasons for selecting this format were to reduce the time spent rewriting publications into thesis chapters, as well as to enhance my writing for publication skills. Permission has been obtained to submit the thesis in alternative format (Appendix 3, page 182).

This PhD thesis contains the following chapters:

Chapter Two defines osteoporosis in children, summarises the factors that affect bone health with a critical review of current techniques for diagnosing vertebral fractures in children, and finally describes the different scoring systems available for recording

severity of vertebral fractures in children. This entire chapter has been published as a review paper in *Pediatric Radiology*. DOI: 10.1007/s00247-018-4279-5

Chapter Three includes an evaluation of the observer reliability and diagnostic accuracy of the semi-automated 6-point technique developed for vertebral fracture diagnosis in adults using a semi-automated software programme (SpineAnalyzer™) on lateral spine images of 137 children and adolescents. This evaluation is the largest to assess vertebral morphometry in children using the semi-automated 6-point technique software. This chapter has been published as an original article in *Clinical Radiology*. DOI.org/10.1016/j.crad.2017.04.010

Chapter Four extends the results of Chapter Three and compares the results of the 6-point technique software (SpineAnalyzer™) on radiographs to a novel 33-point technique programme (AVERT™). Both programmes were applied on radiographs and DXA images performed on the same day on 50 children by two observers. To my knowledge, this is the first report which has aimed to assess two programmes on two different modalities (DXA and radiographs) for the identification of vertebral fractures in children. This chapter will be submitted for publication in *Clinical Radiology*.

Chapter Five further assesses the most effective MXA method identified in Chapter Four - DXA-VFA with AVERT™. This technique was applied on a range of 420 children and adolescents with chronic conditions associated with vertebral fractures. This is the largest study to date assessing MXA for vertebral fracture identification in this population. This chapter will be submitted for publication in *The Lancet Child & Adolescent Health*.

Finally, *Chapter six* includes an integrated summary and discussion of the studies presented in this thesis before drawing its final conclusions and suggesting directions for future work.

Chapter Two
Literature Review

Alqahtani, F.F. & Offiah, A.C. (2018) Diagnosis of osteoporotic vertebral fractures in children. *Pediatr Radiol*, available Online. https://doi.org/10.1007/s00247-018-4279-5

2. Osteoporotic vertebral fracture diagnosis in children: Literature review

Fawaz F Alqahtani[1,2], Amaka C Offiah[1,3]

Correspondence: Mr Fawaz F Alqahtani

Academic Unit of Child Health
Department of Oncology and Metabolism
University of Sheffield, Medical School
Damer Street Building
Sheffield Children's Hospital
Western Bank
Sheffield S10 2TH
Tel: 0114 271 7228
E-mail: ffmalqahtani1@sheffield.ac.uk

[1]   Academic Unit of Child Health, University of Sheffield, Sheffield, UK

[2]   Department of Radiological Sciences, College of Applied Medical Sciences, Najran University, Najran, Kingdom of Saudi Arabia

[3]   Radiology Department, Sheffield Children's NHS Foundation Trust, Sheffield, UK

2.1 Abstract

Osteoporosis is a generalised disorder of the skeleton with reduced bone density and abnormal bone architecture. It increases bone fragility and renders the individual susceptible to fractures. Fractures of the vertebrae are common osteoporotic fractures. Vertebral fractures may result in kyphosis or height loss and because they may be clinically silent, it is imperative that vertebral fractures are diagnosed in children accurately and at an early stage, so that the necessary medical care can be implemented. Traditionally, diagnosis of osteoporotic vertebral fractures has been from lateral spine radiographs, however, a small number of studies have shown that dual energy x-ray absorptiometry (DXA) is comparable to radiographs for identifying vertebral fractures in children, allowing reduced radiation exposure. The diagnosis of vertebral fractures from DXA is termed vertebral fracture assessment (VFA). Existing scoring systems for vertebral fracture assessment in adults have been assessed for use in children, however there is no standardisation and observer reliability is variable. This literature review suggests the need for a semi-automated tool that (compared to the subjective and semi-quantitative methods currently available) will allow more reliable and precise detection of vertebral fractures in children.

2.2 Keywords

Children – Dual energy x-ray absorptiometry – Diagnostic scoring system – Osteoporosis – Vertebral fracture – Vertebral fracture assessment

2.3 Introduction

Fractures are common in childhood. Around one third of United Kingdom children will suffer from at least one fracture during their childhood [1]. Osteoporotic vertebral fractures are increasingly recognised in children with either primary e.g. osteogenesis imperfecta [2] or secondary low bone mineral density including acute lymphoblastic leukaemia, inflammatory bowel disease and glucocorticoid use [3,4]. Nearly one in five children with a rheumatological condition will have a vertebral fracture[5] and rates are similar or even higher in other conditions e.g. 16% in acute lymphoblastic leukaemia[6], up to 75% in Duchenne muscular dystrophy[7] and up to 100% in severe forms of osteogenesis imperfecta (personal experience of the senior author). Outside the context of major trauma, the occurrence of vertebral fractures in children is an indication of pathological bone fragility and precise and early diagnosis is imperative so that appropriate medical care can be initiated.

Techniques used for detecting and analysing vertebral fractures in clinical and/or research practice include conventional radiography, computed tomography (CT), magnetic resonance imaging (MRI) and dual energy x-ray absorptiometry (DXA). Traditionally, the most common method for diagnosing vertebral fractures is x-ray, although DXA has now shown itself able to diagnose vertebral fractures with the advantage of also determining bone mineral density [8]. Vertebral fracture assessment (VFA) is the term given to the diagnosis of vertebral fractures from DXA scans [9]. This technology is more or less in routine clinical use in adults, complimented by validated scoring systems [10,11]. Conversely, VFA is less widely used in children and specific paediatric scoring systems are yet to be fully validated [8,12-15]. Table 2.1 summarises the advantages and disadvantages of DXA and radiographic imaging of the spine.

This review defines osteoporosis in children, summarises the factors that affect bone health, highlights current diagnostic techniques in respect to vertebral fracture diagnosis and outlines the different scoring systems available for recording severity of vertebral fractures in children.

Table 2. 2 Comparative advantages and disadvantages of DXA and radiographic imaging of the spine

	Advantages	Disadvantages
DXA	- Easily and rapidly applicable during BMD measurement - Less cost - Less radiation dose (1–6 μSv) - The entire spine is visualised on a single lateral image - Lateral decubitus and supine position with some scanners with a 'C' arm function e.g. Hologic - Parallax distortion is much less common on VFA images	- Poor visualisation of vertebrae above the level of T7 - Low spatial resolution (but recent advances in VFA technology have improved image resolution)
Radiograph	- Less noisy - High spatial resolution - More vertebrae are evaluable, especially in the upper thoracic spine.	- High radiation dose (232.7 μSv) - The thoracic and lumber spine require two separate images - Only operable in the lateral decubitus position - Parallax distortion are common with cone-beam X-rays

2.4 Osteoporosis

2.4.1 Definition of osteoporosis in children

According to the International Society for Clinical Densitometry (ISCD) Position Statement [16], in children without exposure to high-energy trauma or local disease, detection of one or more vertebral compression (crush) fractures (defined as a 20% reduction in vertebral body height) indicates osteoporosis. This assessment is then

complemented by the determination of bone mineral density (BMD), to give a complete evaluation of bone health. In individuals without vertebral compression (crush) fractures, osteoporosis is diagnosed based on the presence of a BMD Z-score $\leq$ -2.0 as well as a clinically significant fracture history; the latter being defined by one or more of the following:

i.   Two or more long bone fractures by the age of ten years;

ii.   Three or more long bone fractures by the age of 19 years [16].

It is worth noting, therefore, that by definition, a child can neither be diagnosed with osteoporosis solely based on DXA -BMD measurements nor prior to having experienced at least one fracture.

2.4.2 Factors affecting bone health

Peak bone mass, defined as the highest bone mineral content (BMC) reached in an individual's lifetime, plays a major role in determining osteoporotic fracture risk [17]. Bone strength depends on the quantity of bone present; in general, the higher the bone density, the stronger the bone. Bone density has a direct relation to a bone's material properties and ultimate breaking strength [18]. Therefore, fractures at lower level of force more commonly occur with weaker bone which has lower BMD. Bones of children have a lower BMC than normal adult bone. They are also more elastic (less stiff) and so can absorb generally more energy before fracture occurs [19]. Some researchers have suggested that BMD may be lower in some children who sustain fractures [20-22]. This is an important consideration in children who have bone density issues including osteogenesis imperfecta, acute lymphoblastic leukemia, rheumatological conditions, Duchenne muscular dystrophy and glucocorticoid use.

According to Golden et al [23], factors affecting bone health may be classified as modifiable and non-modifiable. Modifiable factors such as diet, calcium, vitamin D, body weight, exercise and puberty/hormonal status are discussed in [24,25] and are beyond the scope of this review, which concentrates on the non-modifiable factors sex, age and ethnicity.

2.4.2.1 Sex

Generally, it has been observed across all ethnicities that in adults, fractures are more likely to occur in females, especially older postmenopausal women. During childhood on the other hand, boys sustain more fractures than girls at all ages. According to the study by Cooper et al, which reviewed the data from 682 general practices in the United Kingdom on the General Practice Research Database [1], approximately 52,624 boys and 31,505 girls had one or more fractures over the 11-year follow-up period, giving a rate of 133.1/10,000 person-years. Fractures were more common in boys (incidence rate, 161.6/10,000 person-years) than in girls (102.9/10,000 person-years). Other authors report that more than 50% of boys and 40% of girls have at least one childhood fracture [26].

Boys attain a higher bone density than girls at both lumber spine and femoral neck, but their peak values are reached at an older age [27]. In a longitudinal study conducted on 266 healthy children (136 males) aged 4-27 years (mean 13 years), total BMD density and lumbar spine and femoral neck BMD were measured [28]. Males had a higher total body BMD, attributed to their greater weight and lean tissue mass. In addition, boys are more likely to go through rapid rates of bone mineral accrual than girls. The reasons for an increase in BMC in boys include that they have longer, wider bones with thicker cortices than girls. Recent analysis of 18-year-old males and age-, height- and

weight-matched females reported that long bone width was found to be greater at the hip and distal tibia as measured by DXA and pQCT, respectively, in boys compared to girls [29]. In other words, males had greater width and bone area and thus had greater BMC, volumetric BMD, and cortical area and thickness at different sites, including spine, hip, and femur and tibia compared with females. However, another paediatric study by Clark et al. [30] measured humeral dimensions on 551 boys and 548 girls, concluding that in children at Tanner stage 1, humeral length was similar in boys and girls, but width (1.92 vs 1.88 cm, P<0.001) and area (47.7 vs 46.9 cm2 , P<0.001) were greater in boys, resulting in a greater aspect ratio of the humerus (7.78 vs 7.53, P<0.001) (aspect ratio of the humerus was calculated as humeral width divided by length). The authors suggested that further research is needed to determine whether humeral aspect ratio is an important factor of biomechanical strength and fracture risk, particularly in adult populations.

Given all of this, it is also still uncertain whether the increase in fracture risk in boys is because of a reduction in bone mass or due to other factors such as alterations in lifestyle [31]. Thandrayen et al. suggest the latter (i.e. that boys are more active than girls, thus increasing their risk of fracture) [32].

2.4.2.2 Age

During childhood and adolescence, BMC and BMD increase significantly; the increase in BMD is associated with an increase in bone size. BMC and BMD continue to rise across multiple skeletal sites of girls of 16 and boys of 17 years old (i.e. even after growth has ceased) [31]. Lu et al. studied the influence of age and growth on the total body, lumbar spine and femoral neck bone mineral density of 209 healthy subjects (52% boys), aged 5-27 years. There was a considerable age-dependent increase in BMD in both sexes at all sites, except the femoral neck in girls, which peaked at age

14 years [28]. This increase achieved the highest level around the age of 17.5 years in boys and 15.8 years in girls.

An additional wide-ranging study on 84,129 subjects showed that the fracture incidence increased in children between the ages of 4 and 17 years in both sexes [28]. A recent paper studied the pattern of fractures in individuals children aged 0-19 years who had been diagnosed with 10,327 fractures confirmed by radiographs [33]. The peak incidence of fractured was at age of 11-12 years in females and 13-14 years in males, where the most common fracture sites were the distal forearm, tibial/fibular shaft and the forearm shaft with 24%, 13% and 11% respectively. Another recent study showed that most of fractures are located in the upper extremities (73%), 22% in lower extremities and less than 5% in axial skeleton including spine [34]. Both studies showed that the incidence of fractures increases with the age, a pattern explained by changes in children's activity patterns over time.

In short, the reasons behind this high incidence of fractures in childhood are not clear. Some studies [35-37] suggest that low bone mass (caused by one or more of the modifiable and/or non-modifiable factors) may contribute.

2.4.2.3 Ethnicity

Various studies have observed differences in bone strength across different ethnic groups. It is believed that Caucasians are more at risk of fracture than Africans and Latinos, with Asians being most at risk (owing to their relatively small bone size) [31,32, 38-40].

Thandrayen et al., in their 2009 study, compared two ethnic groups of South African children, the first being black and the second white children, with both groups being of the same age and having the same sex-related distribution of fractures. They showed that white children had almost double the risk of fractures compared to black and mixed

ancestry children [32]. The reason for this may be explained by the 2007 study by Kalkwarf et al., which was conducted on 6 to 16 year old girls and boys of varying ethnicity [31]. At all ages, the BMC and BMD of the radius, hip and total body were greater for Africans compared to other ethnicities. It was discovered that African girls had more rapid pubertal development. In addition, it was identified that black girls and boys were of increased weight and height. Hammami et al. reported a greater BMC in black new born children compared to white [41]. Another study conducted on 48 healthy infants showed slightly higher whole-body BMC in black infants compared to white infants [42]. In addition, forearm BMC was found to be higher in black compared to white children (131 children aged between 1-6 years old) [43]. It is worth noting that most studies to date have focused on Africans and Caucasians; other ethnic groups should be included in future studies.

2.5 Assessment of bone mass and structure

Low bone mass has traditionally been considered a disease suffered by the elderly, but it is now being diagnosed in children in relatively large numbers [12]. It is generally said that if the skeletal structure of a child is weak, it is very likely to remain weak into adult life [44]. There are several non-invasive imaging techniques to assess the risk of fracture and although not all are in routine clinical use for adults and/or children, each has its own advantages and disadvantages. These methods include DXA, quantitative CT, peripheral quantitative CT, high-resolution peripheral quantitative CT, quantitative ultrasound (QUS) and MRI.

2.5.1 Dual energy x-ray absorptiometry

DXA was first introduced in the late 1980s, mostly for use in postmenopausal female patients, and is now available for common use around the world. The fundamental

principle of DXA is the measurement of X-rays transmission through the body with high and low energy photons. The usage of two different energies allows differentiation between soft tissue and bone as follows: 1) low energy X-rays are attenuated by soft tissue; 2) high energy photons attenuated by both soft tissue and bone. Then, it is possible to measure the amount of bone by subtracting the amount of soft tissue from soft tissue and bone. Bone mass is measured by DXA as BMC (g) or areal BMD (aBMD):

$$aBMD\ (g/cm^2) = BMC\ (g)/\ Bone\ area\ (cm^2)$$

Comparison can then be made with reference values obtained from healthy children of the same age, sex and ethnic background. Results are expressed as a Z-score; that is the number of standard deviations the BMC or BMD deviates from the expected mean. Z-score measurement requires knowing the age, sex and ethnicity of a specific population of the BMC or aBMD to allow direct comparison to those who have the same characteristics in the reference population, as shown in following Eq:

$$Z\text{-score} = (observed\text{-}mean)\ /\ standard\ deviation$$

It is important that Z-score calculations are done to set a reference standard. This should be obtained on 1) data collected from the same machine manufacturer and software, 2) provision sex and ethnicity specific reference curves, and 3) a large sample size.

The development of suitable algorithms allowed for paediatric DXA imaging from the early 1990s. The developed algorithms detect bone/soft tissue interfaces even in children with low bone density [45]. Although the gold standard for the assessment of bone mass and structure in adults is central DXA of the total hip or femoral neck [46], in children, whole body and lumbar spine (L1-L4) DXA are routine and highly reproducible [47].

34

The main advantages of DXA are short scanning times (30-60 sec), low cost, accessibility and relatively low radiation exposure (1–6 µSv) [48].

There are several limitations of DXA in growing children, including inability to account for soft tissue inhomogeneity, inclusion of the posterior elements of the vertebrae in anteroposterior imaging of the spine and dependence of the results on bone size and morphology [49]. DXA calculates areal BMD by dividing the BMC by the bone area without accounting for the depth of bones. Therefore, areal BMD may be falsely elevated in larger bones; in other words, DXA is affected by the actual size of the patient. Finally, independent assessment of cortical and trabecular bone is not possible with DXA [50]. These restrictions of DXA are a hindrance for the assessment of bone density in infants and growing children and adolescents.

To adjust for variations in bone size and to reflect the volumetric BMD, areal BMD may be modified using various mathematical techniques. One technique involves the calculation of bone mineral apparent density, by dividing BMC by bone volume rather than area [51]. Calculation of bone mineral apparent density is relatively reliable for the hip and spine, where the shape of the bone is similar to a cylinder or cube, respectively. Other researchers have suggested that bone area and measures of body size be included in a multiple regression manner that would allow the incorporation of body size in the calculation of BMC [52]. Another mathematical method as suggested by Molgaard et al. [53], involves a three-step evaluation: height for age, bone area for height, and BMC for bone area and allows an assessment as to whether the child has short, narrow or light bone structure.

Although some studies have provided paediatric reference data for children and adolescents of different sex, age and ethnicity [54,55], there is a lack of normative data for toddlers and infants (i.e. less than two years old). The reasons for this are mainly

due to difficulty in positioning this population appropriately, causing movement artefact and technical challenges in measuring their small bones. Unfortunately, common artefacts can be accrued when children are scanned with DXA which may affect BMD accuracy and precision. Artefacts should be limited to only metallic artefacts that cannot be removed such as rods, screw, pacemakers, feeding tubes and surgical clips. External objects such as leg braces, plaster casts or bras should be removed prior to scanning, or the scan should be rescheduled when these devices are no longer required. Other artefacts include movement and positioning aids such as pillows and sandbags. These should also be removed as they can also can affect BMD [56]. The ISCD guidelines recommended omitting any bony region of interest that may affect BMD accuracy and precision. However, removing the artefact function is not available for all scanners. For example, Hologic does not allow the operator to remove artefacts (e.g. deleting the very high-density pixels that are caused by metal), but GE lunar allows them to be painted out. Therefore, technologists and clinicians must be aware of any artefact that may affect the scan's outcome.

Other studies have established DXA reference data for infants and toddlers [57-60], but participant number have been small or limited to specific populations. Kalkwarf et al [61] provided normative bone density data of the lumber spine in 369 children aged between 1 and 36 months, however further studies are needed to demonstrate age, sex and race differences in this population. Finally, availability of DXA may vary from region to region, however this procedure is presently broadly accessible and certainly is the most widely used for bone density measurement in children.

2.5.2 Quantitative computed tomography

Distinct calculations of cortical and trabecular volumetric BMD can be obtained through QCT using a standard CT scanner. QCT measures true volumetric BMD (g/cm^3) independent of body size and images to a resolution of approximately 1000 µm. QCT 1) allows the calculation of bone size and geometry, both of which affect bone strength and 2) can assess volumetric BMD at axial and peripheral sites. However, QCT is not a preferred option for use in children because of the high radiation dose (2.5-3.0 mSv) [48] and the absence of normative paediatric data.

2.5.3 Peripheral quantitative computed tomography

Dedicated CT scanners provide an assessment of bone morphology, volumetric density and three-dimensional images at peripheral sites (e.g. distal radius, distal tibia). They allow independent evaluation of cortical and trabecular bone with less radiation exposure than standard quantitative CT (< 0.003mSv) [48]. Cortical and trabecular measurements can be obtained from a single scan performed either at a specified but variable distance (depending on length of the bone) e.g. the 4% or 8% length of the distal radius or at a fixed site e.g. 10mm from the growth plate [62,63]. There are many advantages to using pQCT scanners in children. These include their small size, the scanner's relatively low cost (around £65K), calculation of volumetric BMD (vBMD), instead of the projected areal BMD obtained by DXA, and low radiation dose. However, reproducibility, positioning and a long acquisition time (3min) that may prompt unwanted movement in the patient remain a problem and limit its use on children [64].

2.5.4 High-resolution peripheral quantitative computed tomography

Similar to pQCT, direct evaluation of bone micro architecture, an accurate measure of volumetric BMD and an estimation of bone strength using finite element analysis are all possible with HR-pQCT. The advantages of HR-pQCT include high image resolution (82 µm) and low radiation exposure (< 0.005 mSv) [50]. However, with limited availability of scanners, this technique is mainly used for research purposes in children. Other disadvantages are 1) the relatively long scanning time (2 to 3 min), which may be problematic for children, leading to movement artefact and 2) the need for a fixed scanning site, which complicates the interpretation of longitudinal studies in growing children.

2.5.5 Quantitative ultrasound

Due to DXA limitations, QUS has been proposed as an alternative technique to measure bone properties in children. Advantages of QUS include its avoidance of ionizing radiation, cost effectiveness and transportability. QUS depends on the attenuation of the ultrasound beam as it passes through a particular region of interest [48]. The latest instruments in the market are upgraded versions capable of giving accurate measurements for BMD, in addition to reflecting parameters of bone quality and strength. Due to the large amount of soft tissue and muscles at axial sites, QUS can only be applied to appendicular bones including phalanges, radius, calcaneus, patella and tibia. The calcaneus is the most commonly measured site, but can be problematic. Many of QUS scanners have fixed transducers that are designed to fit an adult foot; therefore, in children's feet these scanners may not fully capture the correct region of the heel. However, the newest generation of devices have overcome the problem by providing shims to reposition small feet or using portable transducers that can be utilised on the heel. Furthermore, heel size may affect the measurement of QUS

parameters including broadband ultrasound attenuation (BUA) and speed of sound (SOS). A study on 491 healthy children and adolescents showed significant correlation of heel width with BUA ($r= 0.20$, $p<0.005$ in boys; $r= 0.27$, $p<0.05$ in girls) and with SOS ($r=-0.19$, $p<0.005$ in boys; $r=-0.08$, $p<0.05$ in girls) [65]. A similar study showed significant correlations between QUS and age ($r = 0.34-0.54$), height ($r = 0.13-0.56$) and weight ($r = 0.30-0.60$), and between QUS and BMD measurements ($r = 0.44-0.70$) [66]. Therefore, children with small feet should be measured with a smaller ROI diameter than those with larger feet.

In general, studies of QUS have shown good intra and inter operator reliability with coefficient of variation ranging from 0.3% to 3.7% and 0.3% to 1.2% respectively. The reliability of QUS in children and adolescents has been assessed in several studies [67-70] and although it may potentially be useful for bone density assessment, its true utility from a clinical perspective has not been adequately addressed in children and it remains a research tool.

2.5.6 Magnetic resonance imaging

Like QUS, MRI has the advantage of avoiding the use of ionizing radiation. However, there is a lack of reference scanning protocols and normative data for bone density assessment in children and its use is still limited to research studies. Further drawbacks of MRI are the long scanning time (acquisition time of up to 10 -20 min) [71,72], requirement for specialised coils and the environment of the scanner room (isolated from carers/parents, noisy), which may not be child friendly [73]. Despite these disadvantages, some studies have suggested that MRI is a promising technique to evaluate bone properties in children [74-77]. However, prior to widespread clinical use

of MRI to determine bone mass in children, technical and software advancement is required to improve the reproducibility of measurements.

2.6 Imaging of vertebral fractures

Methods that may be used for detecting and analysing vertebral fractures in clinical practice are conventional radiography, DXA, MRI and CT. Differences between these imaging techniques relate to radiation dose, accessibility, cost and patient convenience. These various factors are summarised in Table 2.2.

Table 2.2 Imaging modalities for the detection of vertebral fractures in children

Modality	Spatial Resolution (μm)	Effective Radiation Dose for whole spine scanning (μSv)	Scan Time (min)	Approximate Cost (single scan - including cost of reporting in pound sterling)
Conventional radiography	100×100	233 [78]	<1	37
Computed tomography	1000×1000	10000 [12]	<1	74 - 100
Magnetic resonance imaging	234×234×500	None	10-30	120 - 163
Dual energy x-ray absorptiometry	350×350	3 [69]	<1	58

2.6.1 Conventional radiography

Currently, x-ray imaging is the most common imaging tool for vertebral fracture detection in children and adolescents. It is frequently the initial imaging investigation of choice for back pain when skeletal disease/vertebral pathology is suspected, since the resolution is excellent. However, there is significant radiation exposure (232.7 µSv)[78], equal to 12 months' background radiation [79].

Assessment of the height and shape of the T4-L4 vertebral bodies from lateral thoracic and lumbar spine radiographs is the standard method [9]. Vertebral levels T1 to T3 are not routinely assessed because of the difficulty in their visualisation due to the summation caused by overlying structures such as intrathoracic organs and the patient's shoulders. The normal physiological wedging that may be seen of the mid-thoracic vertebrae (T5 to T7) should not be mistaken for fracture.

The accuracy of vertebral fracture diagnosis in children from conventional radiographs has been evaluated in several studies, summarised in Table 2.3 [4,12,13,78,80,81]. In general, vertebral levels T7–L4 are highly visible (visibility ranging from 88% to 99.8%), whereas visibility is more limited in the upper part of the thoracic spine (T4-T7) for the reasons given above and because of often poor image quality and poor patient positioning. Inter and intraobserver agreement for vertebral readability range between kappa of 0.33 to 0.98 and 0.43 to 0.76 respectively and for fracture diagnosis range between kappa of 0.43 to 0.66 and 0.52 to 0.76 respectively. Unfortunately, despite the radiation dose, conventional radiographs remain the gold standard imaging modality for diagnosing vertebral fractures in children, mainly due to convenience and availability.

Table 2.3 Summary of more recent published studies for vertebral fracture diagnosis in children

Reference	Number of Patients	Median Age (Years)	Scoring System	Imaging Modality	Number of Observers	Sensitivity and Specificity	% Agreement	Agreement Level (Kappa Score)
Mayranpaa et al 2007 [12]	65	12.1	Genant semiquantitative technique	Radiographs and dual energy x-ray absorptiometry	1 Radiologist 1 Orthopaedic Spine Surgeon	Not available	Not available	<u>Radiographs</u> Interobserver 0.98 <u>Dual energy x-ray absorptiometry</u> Interobserver 0.34
Halton et al 2009 [4]	186	5.3	Genant semiquantitative technique	Radiographs	2 Radiologists	Not available	Not available	Interobserver for fracture defined as Grade 1, 2 or 3 0.44 (95% confidence interval 0.28–0.59) Interobserver for fracture defined as Grade 2 and 3 0.66 (95% confidence interval 0.46–0.87)

Siminoski et al 2014 [80]	186	9.6	Genant semiquantitative technique	Radiographs	3 Radiologists	Not available		Readability* Interobserver 1) *Vertebral level* 0.33 to 0.50 2) *Patient level* 0.29 to 0.46 Intraobserver 1) *Vertebral level* 0.43 to 0.64 2) *Patient level* 0.41 to 0.61
							Vertebral fracture: Interobserver 1) *Vertebral level* 96% to 97% 2) *Patient level* 82% to 85% Intraobserver 1) *Vertebral level* 96% to 98% 2) *Patient level* 84% to 92%	Vertebral fracture: Interobserver 1) *Vertebral level* 0.45 to 0.54 2) *Patient level* 0.43 to 0.48 Intraobserver 1) *Vertebral level* 0.52 to 0.72 2) *Patient level* 0.52 to 0.76
Diacinti et al 2015 [13]	58	7.0	Genant semiquantitative technique	Dual energy x-ray absorptiometry	2 Radiologists	Sensitivity 96% Specificity 100 %	Interobserver 1) *Vertebral level* 93.9% 2) *Patient level* 98.3%	Interobserver 1) *Vertebral level* 0.81 (95 % confidence interval 0.76–0.86) 2) *Patient level* 0.96 (95 % confidence interval 0.89–1.03)

Kyriakou et al 2015 [14]	165	13.4	Genant semiquantitative technique	Dual energy x-ray absorptiometry	2 Non-Radiologists	Sensitivity 75% Specificity 98%	<u>Readability</u> Interobserver 1) *Vertebral level* 94% 2) *Patient level* 84% <u>Vertebral fracture:</u> Interobserver 1) *Vertebral level* 99% 2) *Patient level* 91%	<u>Readability</u> Interobserver 1) *Vertebral level* 0.73 (95% confidence interval, 0.68–0.73) 2) *Patient level* 0.66 (95% confidence interval, 0.56, 0.77) <u>Vertebral fracture:</u> Interobserver 1) *Vertebral level* 0.85 (95% confidence interval, 0.79–0.91) 2) *Patient level* 0.78 (95% confidence interval, 0.66–0.87)
Adiotomre et al 2016 [78]	250	11.5	Simplified algorithm-based qualitative technique	Radiographs and dual energy x-ray absorptiometry	3 Radiologists	<u>Dual energy x-ray absorptiometry</u> Sensitivity 70% (95% confidence interval 58–82%) Specificity 97% (95% confidence interval 94–100%) <u>Radiographs</u> Sensitivity 84% (95% confidence interval 0.70–0.99) Specificity 72% (95% confidence interval 0.47–0.97)	<u>Dual energy x-ray absorptiometry</u> Interobserver 66% Intraobserver 86% <u>Radiographs</u> Interobserver 64% Intraobserver 82%	<u>Dual energy x-ray absorptiometry</u> Interobserver 0.37 Intraobserver 0.63 <u>Radiographs</u> Interobserver 0.42 Intraobserver 0.62

Crabtree et al 2017 [15]	80	12.0	Genant semiquantitative technique	Dual energy x-ray absorptiometry	1 Paediatric Radiologist 2 Paediatricians	Sensitivity 1) *Vertebral level* 66% 2) *Patient level* 82% Specificity 1) *Vertebral level* 95% 2) *Patient level* 78%	Interobserver 1) *Vertebral level* 90.4% 2) *Patient level* 80.1%	Interobserver 1) *Vertebral level* 0.63 (95% confidence interval 0.56–0.69) 2) *Patient level* 0.60 (95% confidence interval 0.42–0.77)
Crabtree et al 2017 [15]	80	12.0	Morphometric analysis	Dual energy x-ray absorptiometry	1 Paediatric Radiologist 2 Clinical Scientists 1 Radiographer	Sensitivity 1) *Vertebral level* 79% 2) *Patient level* 43% Specificity 1) *Vertebral level* 71% 2) *Patient level* 97%	Interobserver 1) *Vertebral level* 72% 2) *Patient level* 71.8%	Interobserver 1) *Vertebral level* 0.32 (95% confidence interval 0.26–0.38) 2) *Patient level* 0.41 (95% confidence interval 0.24–0.59)
Alqahtani et al 2017 [81]	137	12.0	Morphometric analysis using SpineAnalyzer™	Radiographs	1 Paediatric Radiologist 2 Radiographers 2 Medical Students	Sensitivity 18% (95% confidence interval 0.14–0.22) Specificity 97% (95% confidence interval 0.97–0.98)	Not available	Interobserver 0.05–0.47 Intra-observer using intraclass correlation coefficient 0.25–0.61

* Readability refers to the rate of radiographs/dual energy x-ray absorptiometry scans that were of sufficient quality to be interpretable and was calculated by dividing the number of readable vertebrae for each vertebral level/patient level by each observer (intraobserver) and/or by all observers (interobserver)

2.6.1.1 Ideal radiographic technique for children

It is essential to have good quality spine radiographs for the precise evaluation of vertebral fractures and associated deformities. In order to generate good quality lateral images the spine should be as parallel as possible to the radiographic table [82]. Generally, a 115 cm Source to Image Distance (SID) should be maintained. For thoracic radiographs the beam is centred at T7, and for lumbar radiographs at L3. With younger children, it is possible to perform the whole spine on a single radiograph where the central beam points to T12. For accurate lateral views, position patients on their left side with flexed knees and hips [13]. Figures 2.1a and b show lateral thoracic and lumbar spine radiographs of an ideally-positioned patient.

2.6.2 Dual energy x-ray absorptiometry

Despite the limitations of DXA for assessing bone density/predicting fracture risk in children as discussed above, DXA is now considered a significant tool for the assessment and monitoring of bone health [83]. The development of machines that allow lateral imaging has expanded the role of DXA beyond assessing bone strength, to include the assessment of bone morphometry/ vertebral fractures diagnosis – termed VFA.

VFA has exciting potential. It is easily and rapidly applicable during BMD measurement, thus enhancing the management of osteoporotic patients [69], obviates the need for spine radiographs and affords point-of-service convenience for the patient, because the imaging is performed at the same visit and at the same time as the DXA for BMD measurement [78,79]. Perhaps the most significant advantage of VFA is the radiation dose savings that it allows; for instance, a dose as low as 3 μSv has been reported for DXA by some authors [84,85]. Another advantage is the ability of DXA

to acquire the whole lateral spine (of larger patients) in a single image; whereas with radiography, the thoracic and lumber spine require two separate images (Figure2.1).
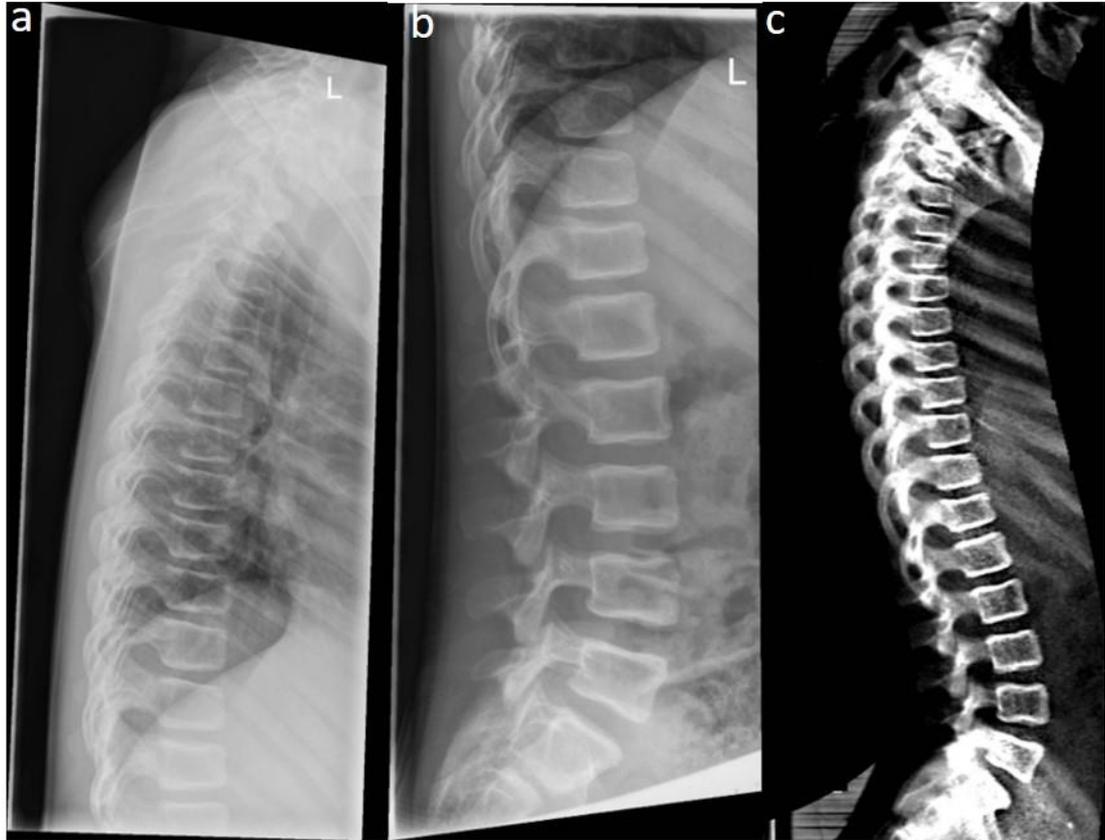
Figure 2.1 An 11-year-old boy with osteogenesis imperfecta. a-c Lateral thoracic (a) and lumbar spine (b) radiographs are juxtaposed to a lateral spine dual energy x-ray absorptiometry scan (c) performed on the same day. The image quality of (c) is non-inferior to (a) and (b), with the advantage of being a single image

VFA images can be obtained either using only one level of X-ray energy (single-energy) or using two X-ray energy levels (dual-energy). With single energy, the images can be acquired more rapidly and have a sharper resolution showing the endplates and cortices better than on dual energy images. However, single energy images have limited visualisation of the vertebrae that is caused by the shadows created by soft tissues, such as the diaphragm area and prominent lung structure. Lunar scanner provides high

quality dual energy images whereas Hologic is provides high quality of single energy images.

Many recent studies have assessed the reliability of VFA in adults and shown good performance with sensitivity and specificity ranging from 62% to 97% and 94% to 99% respectively and observer agreement (kappa score) ranging between 0.24 and 0.98 [86-93].

Mayranpaa et al, in their 2007 study, showed that VFA produces uncertain results in respect to children with low BMD, and they argued that improvements in the image quality of lateral DXA and in scoring systems for VFA were necessary before this approach could be used reliably in children [12]. In contrast, a recent larger study showed similar sensitivity (78% and 72%) and specificity (84% and 72%) for DXA and radiographs respectively, indicating that VFA is as good as conventional radiography for diagnosis of vertebral fractures in children [78]. This study is the only one to address visualisation of non-vertebral body fractures (spondylolysis/spondylolisthesis) and the effects of spinal rods and patient positioning. The study showed that quality of the two modalities was comparable and in fact superior for DXA in the presence of spinal rods[78]. The study by Crabtree et al [15]also demonstrated that DXA is comparable to radiographs for detecting moderate and severe vertebral fractures with sensitivity and specificity of 81.3% and 99.3% respectively . It is worth noting the poor diagnostic accuracy of mild vertebral fractures in children, irrespective of imaging modality, possibly related to poor distinction from normal variants and non-fracture pathology [78].

The latest technological innovation for the new generations of DXA scanners has dramatically improved the detection of vertebral fractures. For example, the high

definition instant vertebral fracture assessment (IVA-HD) imaging tool that is available for Horizon Hologic products has increased the resolution with low radiation dose. Lunar iDXA from GE offers crisp and high-resolution images of spine images through its superior technology including narrow fan beam scanning and multi-view image reconstruction (MVIR) to reduce magnification error. It also uses a unique "K-edge filter" that absorbs the X-rays in the middle energy range and protects the patient against unnecessary exposure.

To summarise, in the past, VFA was found to be inappropriate for paediatric use due to poor image quality, with numerous false positive findings, inability to identify vertebrae in small children and failure to distinguish physiological changes in morphology. However, results of recent studies (Table 2.3) indicate that VFA is a promising technique for diagnosing vertebral fractures in children [12,14,15,78,80].

2.6.2.1 Ideal vertebral fracture assessment technique for children

Depending on the DXA manufacturer, lateral spine scans can be obtained with the patient in a decubitus position (lying on their side). This is the case with GE scanners. In all other cases, the patient remains in a supine position (lying on their back) and rather than the patient moving, the machine's 'c-arm' rotates to obtain a lateral image of the whole spine. This c-arm position is more comfortable for the patient and such scanners are available from Hologic products (e.g. Discovery and Horizon DXA systems).

2.6.2.1.1 Decubitus positioning

The child lies on his/her side on the scanning table. Arms should be kept away from the area to be scanned and should be at a right angle (90º) to the chest. The knees should be flexed upwards towards the chest, so that the spine is parallel to the scanning table. Foam pads may be used to obtain and maintain the required position. Subsequently, the

process of acquiring the DXA scans should be conducted as recommended by the manufacturer. The child should be reminded to stay still throughout the examination [94]. The lateral scan should image the vertebrae from L4 to T4 by determining the starting position of the scan by positioning the laser spot 1 cm above the iliac crest.

2.6.2.1.2 Supine positioning

The child should lie on the scanning table in the supine position with his/her arms raised above their head. The patient's spine should be positioned in the centre of the scanning area and both knees should be raised upwards at 90° using foam pads to straighten up the base of the spine, to allow the spine to be pressed flat to the table to reduce the lumber lordosis.

2.6.3 Computed tomography and Magnetic resonance imaging

CT and MRI are variably used for diagnosis of vertebral fracture in children. CT is the most reliable and accurate method for vertebral morphology evaluation when an acute traumatic fracture is suspected. Disadvantages include a high radiation dose penalty (approximately 10000 μSv for whole spine scanning, equal to 3 years' background radiation) [12], generally reduced availability of machines and relative expense.

MRI has the advantage of not utilising ionising radiation and helps to differentiate the underlying cause of vertebral fractures other than osteoporosis (particularly malignancy), but such scans take a relatively long time and are more costly than other modalities.

Neither CT nor MRI are currently indicated for routine monitoring/diagnosis of osteoporotic vertebral fractures in children.

2.6.4 Biplanar x-ray imaging

The biplanar x-ray imaging system (EOS imaging, Paris, France) is a relatively new imaging solution meeting the specific needs of musculoskeletal imaging [95]. The system produces high-quality images of the whole body, including the whole spine, at lower radiation dose than radiographs (for lateral spine, mean entrance surface dose was 0.37 mGy compared with 2.03 mGy for radiographs) and it has the ability to generate 3D images from simultaneous anteroposterior and lateral 2D images of the whole body [96]. EOS scanning time ranges from 10 to 20 s for a full body scan and 4–6 s for the spine (depending on the patient's height). The EOS system allows imaging in erect position (upright weight-bearing or seated/squatting position), and can image the full length of the body (up to 175 cm). This ability aids physicians in identifying structural spinal pathologies, intervertebral disks, postural dysfunctions and joints and ligaments from multiple angles under normal weight-bearing conditions. However, the lack of spatial resolution compared with the conventional system and wavy images with patients who unable to stand or sit steadily during the scan (i.e. those with neurologic or neuromuscular conditions) are major drawbacks of EOS system [97].

EOS plays a major role in the diagnosis and follow-up of patients with adolescent idiopathic scoliosis[98,99] and has been used to diagnose vertebral fractures in 200 patients aged above 50 years in which it was compared to VFA (DXA device; QDR 4500, Hologic, Bedford, MA)[99]. The sensitivity, specificity and negative predictive values for EOS were 79.7%, 91.6%, and 99%, respectively. Interobserver agreement between two independent readers was very good for EOS (kappa score =0.89), higher than for VFA (kappa score=0.67). We are not aware of any study that has compared EOS with radiographs and/or VFA for the diagnosis of vertebral fractures in children; however, the results in adults suggest that EOS is a good diagnostic tool for the

diagnosis of vertebral fractures. Therefore, it is also likely to be beneficial in children with advantages of high image quality, low dose and rapid acquisition time. Further research studies are needed to assess diagnostic accuracy of vertebral fracture in children using EOS.

2.7 Scoring systems for vertebral fractures

When using any classification system, the normal slight curve of the lower lumbar vertebrae and anatomical changes such as wedging of mid-thoracic and thoracolumbar vertebrae should be borne in mind. Baseline and serial radiographs should be compared with one another in order to document improvement/deterioration in prevalent (i.e. previously identified) vertebral fractures and to detect incident (i.e. new) vertebral fractures [100].

Quantitative morphometric definitions [101-104] and semiquantitative (SQ) assessments of vertebral fractures, including methods by Smith [105] and Kleerekoper [106] have been introduced for the adult population, however this review focuses only on methods that have been used in children.

2.7.1 Subjective visual assessment

The most extensively employed method for the assessment of vertebral fractures is visual assessment of radiographs [107]. Qualitative visual assessment is a helpful method when performed by experts capable of disregarding abnormal appearances that have nothing to do with the osteoporotic fracture. However, due to the subjectivity of the technique, observer reliability is low with inter and intraobserver kappa scores of 0.47 and 0.62 respectively [108]. In other words, the findings of visual assessment greatly depend on the competency of the reader. For this reason, visual assessment is not recommended for epidemiological studies or therapeutic trials.

2.7.2 Genant's semi-quantitative assessment

This SQ grading system was developed in 1993 based on independent analysis of the spine radiographs of 57 postmenopausal women (age 65- 75 years) by three observers [10]. Assessment is made of vertebral shape (crush, concave or wedge) and reduction in posterior, middle and/or anterior vertebral height. As illustrated in Figure 2.2, Grade 0 indicates no fracture (normal) with a height reduction of less than 20%, Grade 1 indicates a minimal fracture with a height reduction in the range of 20-25%, Grade 2 indicates a moderate fracture with a height reduction in the range of 25-40% and Grade 3 indicates a severe fracture with a height reduction of above 40%. Limitations of this study may apply to many important visual characteristics including endplate deformities, short vertebral height and normal variants with some degree of biconcavity (e.g. Cupid's bow).

Although this method was developed for and is the standard tool for diagnosing vertebral fractures in adults, researchers have begun to assess its use for diagnosing vertebral fractures in children [4,13,14,80]. These paediatric studies demonstrate inter- and intraobserver reliability ranging from kappa score = 0.29 to 0.98 and 0.41 to 0.63 respectively and sensitivity and specificity ranging from 66% to 95% and 78% to 100 % respectively. The variability between the different studies for observer reliability and sensitivity/specificity may reflect limitations of the quantitative method such as false positive identification of non-fracture deformities, disparity in fracture prevalence and severity within the study cohorts and misdiagnosis of mild endplate fractures (i.e. mild height loss may be physiological rather than pathological).
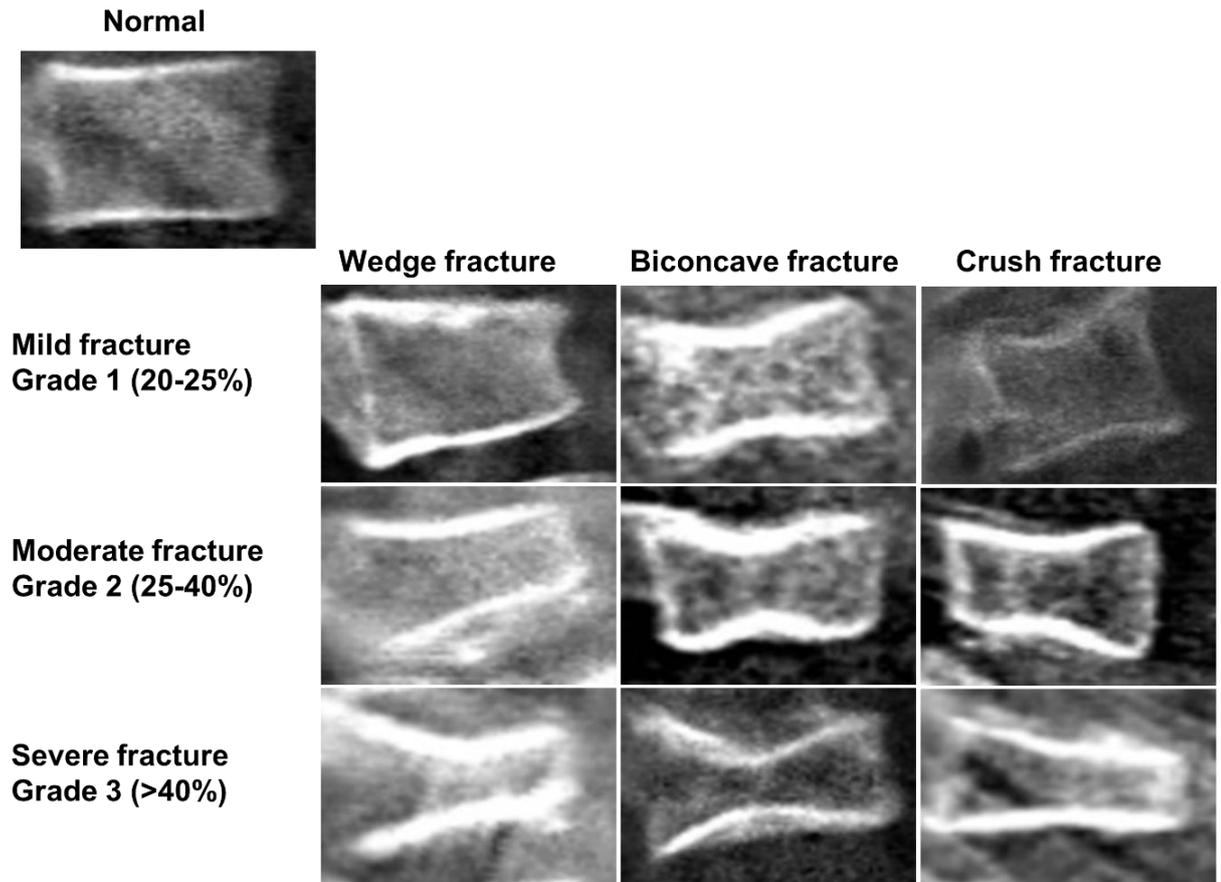
Figure 2.2 Selected lateral spine dual energy x-ray absorptiometry scans from a series of patients demonstrating the semiquantitative visual grading system of Genant et al. [10]

The benefits of Genant's SQ method include its convenience, being less complicated than quantitative methods, improved consistency when compared to qualitative methods and the fact that it can be used by both experts and beginners with acceptable reproducibility and precision [109,110]. However, although reduced compared to subjective visual assessment, the experience and competency of the reader still greatly influences its implementation. A further drawback of this method is that deformation of shape is not taken into consideration while making the evaluation [100]. The fracture is detected by observing the reduction in vertebral height or area but radiological features related to vertebral endplate abnormality are not considered [11].

2.7.3 Algorithm-based qualitative method

The algorithm-based qualitative (ABQ) method is not based on reduced vertebral height alone; this system provides clear guidelines for the evaluation of alterations in the vertebral endplates, helpful in detecting osteoporotic fractures in adults [111]. It has been pointed out that although three grades of severity are present, just as in Genant's system, there is no lower limit for Grade 1 fractures (Grade 1 $\leq$ 25%, Grade 2 > 25%, and Grade 3 > 40%)[11]. Nevertheless, the ABQ approach offers the advantage of addressing potential sources of false-positive detection of vertebral fracture such as "deep inferior" and "step-like" endplates [111].The algorithm serves as a basic guideline for qualitative identification and differentiation of vertebral fracture, non-fracture deformity and normal variant. However, observers need to be fully trained in the application of the method, and the algorithm should be applied with recourse to reference notes on differential diagnoses.

A recent study was the first to analyse the clinical utility of an adapted version of the ABQ approach from radiographs and DXA in children [8]. Development of the scoring system was in two phases; modification of ABQ and simplification of the modified ABQ system (Table 2.4). The researchers showed slight to good inter-observer agreement in 50 patients by both modified ABQ (kappa score= 0·27 to 0·49) and simplified ABQ (kappa score= 0.31 to 45) and moderate intraobserver for simplified ABQ (kappa score= 0.45 to 56). All observers subjectively found simplified ABQ easier and less time-consuming (main reason to use in this thesis), which makes it more appealing for clinical and research use compared to modified ABQ for scoring vertebral fractures in children [8].

Although the ABQ technique is promising as a semi-objective means of classifying osteoporotic vertebral fractures in adults, currently there is only limited research into

this technique in children. Further studies are required to assess whether the simplified

ABQ method is sufficiently reliable to identify and differentiate fracture from non-

fracture deformities in children.

Table 2.4 Modified, simplified and original algorithm-based qualitative grading systems (ABQ) [8].

ABQ	Modified ABQ	Simplified ABQ
Grading scale:	Grading scale: 0 = Normal	Grading scale (focusing on endplates):
1 = Osteoporotic fracture	1a = Vertebral height loss ≤25% (mild #)	Height
2 = Non-osteoporotic short vertebral height	1b = Vertebral height loss >25%-≤40% (moderate #)	0 = Normal 1 = Height loss ≤ 24%
3 = Normal	1c = Vertebral height loss >40% (severe #)	2 = Height loss ≥25%
4 = Uncertain (possible osteoporotic fracture, but uncertain due to atypical appearance or poor image quality)	2 = Non-osteoporotic deformity (please add comment)	Endplates a = Normal
5 = Unable to evaluate (poor image quality or not imaged)	3 = Uncertain or unable to determine due to quality (please add comment)	b = Single endplate affected c = Both endplates affected
Fractures classified on vertebral height reduction:	For 1a, 1b, 1c	Others requiring comments
Mild ≤25%	Shape: Concave/Wedge/Crush (V/W/K)	3 = Non-osteoporotic deformity
Moderate >25% - <40%	Affected end plate: Superior/Inferior/Both (S/I/B)	4 = Uncertain or unable to determine due to quality
Severe ≥40%	Position: Anterior/Middle/ Posterior/Entire vertebral body (A/M/P/E)	

2.7.4 Koerber's technique

Recently, a new scoring system was introduced for assessing spine morphology in children with osteogenesis imperfecta (developed using 268 lateral spine radiographs of 95 patients) [112]. The assessment is based on three criteria: vertebral compression, thoracolumbar kyphosis and deformity type, with a scale of 1 to 5 to defining severity (1= no need for therapy; 5= extreme severe). To record all possible combinations of the three parameters, the authors developed a more detailed severity score system ranging from 1 to 138. The authors state that this evaluation will provide benefit for patients in clinics; however, it seems that this method is limited by being relatively time-consuming (a trained reader needs from 5 to 8 minutes to define the category and severity scores) [112].

2.7.5 Semi-automated techniques

Semi-automated quantitative vertebral morphometry techniques typically employ model-based shape recognition technology to define the shape of all vertebrae between T4 and L4 inclusive (Figure 2.3) [113].

Several methods have been investigated to segment vertebrae based on statistical models [114-117] and can be placed on a spectrum according to their complexity. At one end of that spectrum we have pure edge detector methods based on finding points of high gradient. Methods like Canny et al.'s [118] (which finds extended sets of high gradients using hysteresis thresholding) can be included in this class. These methods provide significant contamination from other bone edges and so require some post-processing. The next point on the spectrum would be methods including level sets such as the Mumford Shah functional [119] and snakes/active contours [120], which make some basic assumptions about the behaviour of the edge (e.g. in snakes, the assumption is that the curvature of the edge is low). At the other end of the

spectrum, we find full statistical shape models (SSMs) like the point distribution models that are used in the hierarchical clustering-based segmentation (HCS), the Random Forest Regression Voting Constrained Local Model (RFRV-CLM) and other appearance modelling techniques. These do not make assumptions about the shape or properties of edges (other than a very basic one that they can be identified by local image features such as a high intensity gradient). Instead, they learn the distribution of shapes from the training data. SSMs can be fitted to images directly by optimising over their parameters using a cost function that maximises the image gradient under the shape. All of the active appearance models have been developed at Manchester since the 1990's, including the RFRV-CLM are improvements on this basic idea, where the cost function could be replaced with one based on image intensity features. The basic idea is to fit the SSM to the image.

For semi-automated programmes, the procedure (termed 6-point morphometric analysis) begins with a manual indication of the estimated centre of each vertebra from T4 to L4. The software then mechanically identifies and marks the standard positions for six-point morphometry measurements. The operator may move these points with the help of the software for improved accuracy. From these six points, anterior (Ha), middle (Hm), and posterior (Hp) vertebral heights are automatically determined by the software. Then, the (Ha: Hp), (Hm: Hp), (Hp: Hp+1) and (Hp: Hp-1) height ratios are calculated (+1 and –1 indicate the vertebrae immediately above and below the vertebra of interest). Each vertebral body is then classified according to its height ratios based on the Genant classification (Grades 0 to 3).

Image
Not for diagnostic use.
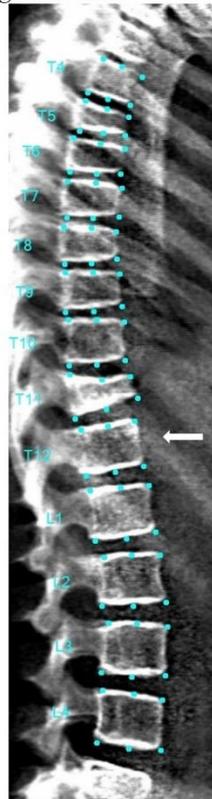
Deformity results

Level	Morphometry (% deformation)			Grade	SQ Score
	Wedge	Bicon.	Crush		
T4	17.3%	5.9%	-65.0%	Normal	-
T5	38.6%	19.7%	18.4%	Moderate	-
T6	26.7%	11.0%	4.5%	Moderate	-
T7	22.1%	1.5%	-1.4%	Mild	-
T8	23.8%	2.5%	8.6%	Mild	-
T9	19.2%	2.2%	12.9%	Normal	-
T10	16.3%	1.7%	-18.9%	Normal	-
T11	50.5%	32.6%	3.2%	Severe	-
T12	1.1%	-0.4%	2.0%	Normal	-
L1	8.2%	-1.0%	-4.5%	Normal	-
L2	2.0%	-3.6%	2.1%	Normal	-
L3	2.5%	-7.3%	2.5%	Normal	-
L4	4.5%	-4.0%	-4.7%	Normal	-

0 - Normal < 20%	1 - Mild 20%-<25%	2 - Moderate 25%-<40%	3 – Severe ≥ 40%

Level	Differential diagnosis
T4	-
T5	-
T6	-
T7	-
T8	-
T9	-
T10	-
T11	-
T12	-
L1	-
L2	-
L3	-
L4	-

Figure 2.3 A 14-year-old girl with osteogenesis imperfecta. Lateral spine dual energy x -ray absorptiometry scan illustrates positioning of points used to outline the vertebral bodies between T4 and L4 using the SpineAnalyzer™ programme. SpineAnalyzer™ identified a severe fracture at T11, moderate fractures at T5 and T6 and mild fractures at T7 and T8. The arrow points to the T12 vertebral body (lowest vertebral body associated with a rib)

SpineAnalyzer™ software (Optasia Medical Ltd, Cheadle, UK) provides a quick and easy method for identifying and reporting vertebral deformities from radiographs and other x-ray-based technology. A recent study concluded that SpineAnalyzer™ is a reliable and ideal system for measuring vertebral height and identifying vertebral fracture from DXA scans in adults ,with significant observer agreement (ranging from 96% to 98.6%) using the Genant SQ method [121]. Although studies have used the semi-automated quantitative software to diagnose fractures in adults, as far as we are aware, only two studies have used this semi-automated six-point software technique to diagnose vertebral fractures in children [15,81].

The study by Alqahtani et al [81] is currently the largest morphometric analysis study in children using SpineAnalyzer™; the study assessed 137 lateral spine radiographs of children aged between 5 and 15 years . Inter and intra-observer agreement, overall sensitivity and specificity are shown in Table 2.2. Another study by Crabtree et al [15] demonstrated poor observer agreement of morphometric analysis for vertebral fracture diagnosis in 80 children (Table 2.2).

The results of these two papers suggest that the diagnostic accuracy of semi-automated systems/ morphometric analysis is not sufficiently high to allow their routine clinical use in children. Therefore, training of current software programmes on paediatric images or development of paediatric specific software and reference values is required.

2.8 Summary

Identification of vertebral fracture is central to the diagnosis of osteoporosis in children. Imaging methods used for detecting and analysing vertebral fractures in clinical and/or research practice include conventional radiographs, DXA (VFA), CT and MRI. While VFA is routine in adults, identification of vertebral fracture in children is still mostly from radiographs. However, recent paediatric studies have shown that DXA VFA has similar sensitivity and specificity to radiographs with lower radiation dose; therefore, DXA should be considered for vertebral fracture diagnosis in children, when feasible. It is likely that EOS will have an increasing role.

There is no agreed standardised method for diagnosing vertebral fracture in children and it is difficult to be certain of the validity of mild fractures.

2.9 Conclusion

There is currently no reliable method of VFA in children. This situation may be improved by the development of a software tool for semi-automated VFA. Such a tool

should be specifically designed for paediatric use and encompass normal physiological variation, which almost certainly accounts for some observer variability in diagnosing vertebral fractures in this population.

2.10 References

1. Cooper C, Dennison EM, Leufkens HGM, Bishop N, van Staa TP (2004) Epidemiology of childhood fractures in Britain: A study using the General Practice Research Database. J Bone Miner Res 19:1976-1981 doi:10.1359/jbmr.040902

2. Bishop N (2010) Characterising and treating osteogenesis imperfecta. Early Hum Dev 86:743-746. https://doi.org/10.1016/j.earlhumdev.2010.08.002

3. Mäkitie O (2013) Causes, mechanisms and management of paediatric osteoporosis. Nat Rev Rheumatol 9:465. doi:10.1038/nrrheum.2013.45

4. Halton J, Gaboury I, Grant R et al (2009) Advanced Vertebral Fracture Among Newly Diagnosed Children With Acute Lymphoblastic Leukemia: Results of the Canadian Steroid-Associated Osteoporosis in the Pediatric Population (STOPP) Research Program. J Bone Miner Res 24:1326-1334. doi:10.1359/jbmr.090202

5. Nakhla M, Scuccimarri R, Duffy KNW et al (2009) Prevalence of vertebral fractures in children with chronic rheumatic diseases at risk for osteopenia. J Pediatr 154:438-443

6. Alos N, Grant R, Ramsay T et al (2012) High incidence of vertebral fractures in children with acute lymphoblastic leukemia 12 months after the initiation of therapy. J Clin Oncol 30:2760

7. Bothwell J, Gordon K, Dooley J, Cummings E, Salisbury S (2003) Vertebral fractures in boys with Duchenne muscular dystrophy. Clin Pediatr (Phila) 42:353-356

8. Adiotomre E, Summers L, Allison A et al (2016) Diagnosis of vertebral fractures in children: is a simplified algorithm-based qualitative technique reliable? Pediatr Radiol 46:680-688. doi:10.1007/s00247-015-3537-z

9. Maricic M (2014) Use of DXA-Based Technology for Detection and Assessment of Risk of Vertebral Fracture in Rheumatology Practice. Curr Rheumatol Rep 16. doi:10.1007/s11926-014-0436-5

10. Genant HK, Wu CY, Vankuijk C, Nevitt MC (1993) Vertebral fracture assessment using a semiquantitative technique. J Bone Miner Res 8:1137-1148

11. Jiang G, Eastell R, Barrington NA, Ferrar L (2004) Comparison of methods for the visual identification of prevalent vertebral fracture in osteoporosis. Osteoporos Int 15:887-896. doi:10.1007/s00198-004-1626-1

12. Mayranpaa MK, Helenius I, Valta H, Mayranpaa MI, Toiviainen-Salo S, Makitie O (2007) Bone densitometry in the diagnosis of vertebral fractures in children: Accuracy of vertebral fracture assessment. Bone 41:353-359. doi:10.1016/j.bone.2007.05.012

13. Diacinti D, Pisani D, D'Avanzo M et al (2015) Reliability of Vertebral Fractures Assessment (VFA) in Children with Osteogenesis Imperfecta. Calcif Tissue Int 96:307-312. doi:10.1007/s00223-015-9960-1

14. Kyriakou A, Shepherd S, Mason A, Ahmed SF (2015) A critical appraisal of vertebral fracture assessment in paediatrics. Bone 81:255-259. doi:10.1016/j.bone.2015.07.032

15. Crabtree NJ, Chapman S, Högler W, Hodgson K, Chapman D, Bebbington N, Shaw NJ (2017) Vertebral fractures assessment in children: Evaluation of DXA imaging versus conventional spine radiography. Bone 97:168-174. https://doi.org/10.1016/j.bone.2017.01.006

16. Bishop N, Arundel P, Clark E et al (2014) Fracture Prediction and the Definition of Osteoporosis in Children and Adolescents: The ISCD 2013 Pediatric Official Positions. J Clin Densitom 17:275-280. doi:10.1016/j.jocd.2014.01.004

17. Bonjour JP, Theintz G, Law F, Slosman D, Rizzoli R (1994) Peak bone mass. Osteoporos Int 4:7-13

18. Hui, S.L., Slemenda, C.W. and Johnston, C.C., (1988). Age and bone mass as predictors of fracture in a prospective study. J Clin Invest 81:1804-1809

19. Currey, J. D., and G. Butler. (1975) The mechanical properties of bone tissue in children. J Bone Joint Surg. 57: 810-814

20. Chan, G.M., Hess, M., Hollis, J. and Book, L.S., (1984). Bone mineral status in childhood accidental fractures. Am J Dis Child, 138:569-570

21. Davie, M.W. and Haddaway, M.J., (1994). Bone mineral content and density in healthy subjects and in osteogenesis imperfecta. Arch Dis Child 70:331-334

22. Henderson, R.C., Lin, P.P. and Greene, W.B., (1995). Bone-mineral density in children and adolescents who have spastic cerebral palsy. JBJS, 77:1671-1681

23. Golden NH, Abrams SA, Comm N (2014) Optimizing Bone Health in Children and Adolescents. Pediatrics 134:E1229-E1243. doi:10.1542/peds.2014-2173

24. Ondrak KS, Morgan DW (2007) Physical activity, calcium intake and bone health in children and adolescents. Sports Med 37:587-600

25. Leeuwen J, Koes B, Paulis W, Middelkoop M (2017) Differences in bone mineral density between normal-weight children and children with overweight and obesity: a systematic review and meta-analysis. Obes Rev 18:526-546

26. Amin S, Melton LJ, Achenbach SJ et al (2013) A distal forearm fracture in childhood is associated with an increased risk for future fragility fractures in adult men, but not women. J Bone Miner Res 28:1751-1759. doi:10.1002/jbmr.1914

27. Bianchi ML, Sawyer AJ, Bachrach LK (2016) Rationale for Bone Health Assessment in Childhood and Adolescence. In: Bone Health Assessment in Pediatrics. Springer, pp 1-21

28. Lu PW, Briody JN, Ogle GD et al (1994) Bone-mineral density of total-body, spine, and femoral-neck in children and young-adults - a cross-sectional and longitudinal-study. J Bone Miner Res 9:1451-1458

29. Nieves, J.W., Formica, C., Ruffing, J., Zion, M., Garrett, P., Lindsay, R. and Cosman, F., (2005). Males have larger skeletal size and bone mass than females, despite comparable body size. J Bone Miner Res 20:529-535

30. Clark, E.M., Ness, A.R. and Tobias, J.H., (2007). Gender differences in the ratio between humerus width and length are established prior to puberty. Osteoporos Int,18:463-470

31. Kalkwarf HJ, Zemel BS, Gilsanz V et al (2007) The bone mineral density in childhood study: Bone mineral content and density according to age, sex, and race. J Clin Endocrinol Metab 92:2087-2099. doi:10.1210/jc.2006-2553

32. Thandrayen K, Norris SA, Pettifor JM (2009) Fracture rates in urban South African children of different ethnic origins: The Birth to Twenty Cohort. Osteoporos Int 20:47-52. doi:10.1007/s00198-008-0627-x

33. Hedström, E.M., Svensson, O., Bergström, U. and Michno, P., (2010) Epidemiology of fractures in children and adolescents: Increased incidence over the past decade: a population-based study from northern Sweden. Acta orthopaedica, 81: 148-153

34. Mäyränpää, M.K., Mäkitie, O. and Kallio, P.E., (2010) Decreasing incidence and changing pattern of childhood fractures: A population-based study. J Bone Miner Res, *25*:2752-2759

35. Clark E, Tobias J, Ness A (2006) Association between bone density and fractures in children: a systematic review and meta-analysis. Pediatrics 117:e291-e297

36. . Ma D, Jones G (2003) The association between bone mineral density, metacarpal morphometry, and upper limb fractures in children: a population-based case-control study. J Clin Endocrinol Metab 88:1486-1491

37. Schalamon J, Singer G, Schwantzer G, Nietosvaara Y (2004) Quantitative ultrasound assessment in children with fractures. J Bone Miner Res 19:1276-1279

38. Cummings SR, Melton LJ (2002) Epidemiology and outcomes of osteoporotic fractures. Lancet 359:1761-1767. doi:10.1016/s0140-6736(02)08657-9

39. Bachrach LK, Hastie T, Wang M-C, Narasimhan B, Marcus R (1999) Bone Mineral Acquisition in Healthy Asian, Hispanic, Black, and Caucasian Youth: A Longitudinal Study1. J Clin Endocrinol Metab 84:4702-4712. doi:10.1210/jcem.84.12.6182

40. Li J-Y, Specker BL, Ho ML, Tsang RC (1989) Bone mineral content in black and white children 1 to 6 years of age: early appearance of race and sex differences. Am J Dis Child 143:1346-1349

41. Hammami, M., Koo, W.W. and Hockman, E.M., (2003) Body composition of neonates from fan beam dual energy X-ray absorptiometry measurement. JPEN J Parenter Enteral Nutr, *27*:423-426

42. Rupich, R.C., Specker, B.L., Lieuw-A-Fa, M. and Ho, M., (1996) Gender and race differences in bone mass during infancy. Calcif. Tissue Int., *58*:395-397.

43. Li, J.Y., Specker, B.L., Ho, M.L. and Tsang, R.C., (1989) Bone mineral content in black and white children 1 to 6 years of age: early appearance of race and sex differences. Am J Dis Childn, 143:1346-1349

44. Ma NS, Gordon CM (2012) Pediatric Osteoporosis: Where Are We Now? J Pediatr 161:983-990. doi:10.1016/j.jpeds.2012.07.057

45. Specker B, Schoenau E (2006) Quantitative bone analysis in children: Current methods and recommendations - Reply. J Pediatr 149:431-432. doi:10.1016/j.jpeds.2006.04.007

46. Mauck KF, Clarke BL (2006) Diagnosis, screening, prevention, and treatment of osteoporosis. Mayo Clin Proc 81:662-672

47. Crabtree NJ, Arabi A, Bachrach LK et al (2014) Dual-Energy X-Ray Absorptiometry Interpretation and Reporting in Children and Adolescents: The Revised 2013 ISCD Pediatric Official Positions. J Clin Densitom 17:225-242. doi:10.1016/j.jocd.2014.01.003

48. Adams JE (2016) Bone Densitometry in Children. Seminars in Musculoskeletal Radiology 20:254-268. doi:10.1055/s-0036-1592369

49. Wren TA, Gilsanz V (2006) Assessing bone mass in children and adolescents. Current osteoporosis reports 4:153-158

50. Digby M, Bishop N, Paggiosi M, Offiah A (2016) HR-pQCT: a non-invasive 'biopsy' to assess bone structure and strength. Arch Dis Child Educ Pract Ed. 101:268-270

51. Carter DR, Bouxsein ML, Marcus R (1992) New approaches for interpreting projected bone densitometry data. J Bone Miner Res 7:137-145

52. Prentice A, Parsons TJ, Cole TJ (1994) Uncritical use of bone-mineral density in absorptiometry may lead to size-related artifacts in the identification of bone-mineral determinants. Am J Clin Nutr 60:837-842

53. Molgaard C, Thomsen BL, Prentice A, Cole TJ, Michaelsen KF (1997) Whole body bone mineral content in healthy children and adolescents. Arch Dis Child 76:9-15

54. Ward KA, Ashby RL, Roberts SA, Adams JE, Mughal MZ (2007) UK reference data for the Hologic QDR Discovery dual-energy x ray absorptiometry scanner in

healthy children and young adults aged 6-17 years. Arch Dis Child 92:53-59. doi:10.1136/adc.2006.097642

55. Crabtree NJ, Shaw NJ, Bishop NJ et al (2017) Amalgamated Reference Data for Size-Adjusted Bone Densitometry Measurements in 3598 Children and Young Adults-the ALPHABET Study. J Bone Miner Res 32:172-180. doi:10.1002/jbmr.2935

56. Morgan, S.L. and Prater, G.L., (2017) Quality in dual-energy X-ray absorptiometry scans. Bone, 104:13-28

57. Specker BL, Beck A, Kalkwarf H, Ho M (1997) Randomized trial of varying mineral intake on total body bone mineral accretion during the first year of life. Pediatrics 99:e12-e12

58. Butte NF, Hopkinson JM, Wong WW, Smith EOB, Ellis KJ (2000) Body composition during the first 2 years of life: an updated reference. Pediatr Res 47:578

59. Ay L, Jaddoe VW, Hofman A, Moll HA, Raat H, Steegers EA, Hokken-Koelega A (2011) Foetal and postnatal growth and bone mass at 6 months: the Generation R Study. Clin Endocrinol (Oxf) 74:181-190

60. Gallo S, Vanstone CA, Weiler HA (2012) Normative data for bone mass in healthy term infants from birth to 1 year of age. Journal of osteoporosis 2012. http://dx.doi.org/10.1155/2012/672403

61. Kalkwarf HJ, Zemel BS, Yolton K, Heubi JE (2013) Bone mineral content and density of the lumbar spine of infants and toddlers: influence of age, sex, race, growth, and human milk feeding. J Bone Miner Res 28:206-212

62. Neu CM, Manz F, Rauch F, Merkel A, Schoenau E (2001) Bone densities and bone size at the distal radius in healthy children and adolescents: A study using peripheral quantitative computed tomography. Bone 28:227-232. doi:10.1016/s8756-3282(00)00429-4

63. Wang QJ, Alen M, Nicholson P et al (2005) Growth patterns at distal radius and tibial shaft in pubertal girls: a 2-year longitudinal study. J Bone Miner Res 20:954-961. doi:10.1359/jbmr.050110

64. Saraff V, Hoegler W (2015) Osteoporosis in children: diagnosis and management. Eur J Endocrinol 173:R185-R197. doi:10.1530/eje-14-0865

65. Van den Bergh, J.P.W., Noordam, C., et al., (2000) Calcaneal ultrasound imaging in healthy children and adolescents: relation of the ultrasound parameters BUA and SOS to age, body weight, height, foot dimensions and pubertal stage. Osteoporos Int , 11:967-976

66. Sundberg, M., Gärdsell, P., Johnell, O., Ornstein, E. and Sernbo, I., (1998) Comparison of quantitative ultrasound measurements in calcaneus with DXA and SXA at other skeletal sites: a population-based study on 280 children aged 11–16 years. Osteoporos Int, 8:410-417.

67. Pereira-da-Silva L, Costa A, Pereira L et al (2011) Early high calcium and phosphorus intake by parenteral nutrition prevents short-term bone strength decline in preterm infants. J Pediatr Gastroenterol Nutr 52:203-209

68. Tomlinson C, McDevitt H, Ahmed S, White M (2006) Longitudinal changes in bone health as assessed by the speed of sound in very low birth weight preterm infants. J Pediatr 148:450-455

69. Roggero P, Giannì ML, Orsi A et al (2007) Postnatal "speed of sound" decline in preterm infants: an exploratory study. J Pediatr Gastroenterol Nutr 45:615-617

70. Wang KC, Wang KC, Amirabadi A, Cheung E, Uleryk E, Moineddin R, Doria AS (2014) Evidence-based outcomes on diagnostic accuracy of quantitative ultrasound for assessment of pediatric osteoporosis—a systematic review. Pediatr Radiol 44:1573-1587

71. Krug R, Burghardt AJ, Majumdar S, Link TM (2010) High-resolution imaging techniques for the assessment of osteoporosis. Radiologic Clinics 48:601-621

72. Link T, Majumdar S, Grampp S et al (1999) Imaging of trabecular bone structure in osteoporosis. Eur Radiol 9:1781-1788

73. Ward KA, Link TM, Adams JE (2016) Tools for measuring bone in children and adolescents. In: Bone Health Assessment in Pediatrics. Springer, pp 23-52

74. Bass SL, Saxon L, Daly R, Turner CH, Robling AG, Seeman E, Stuckey S (2002) The effect of mechanical loading on the size and shape of bone in pre-, peri-, and postpubertal girls: a study in tennis players. J Bone Miner Res 17:2274-2280

75. McKay HA, Sievänen H, Petit MA et al (2004) Application of magnetic resonance imaging to evaluation of femoral neck structure in growing girls. J Clin Densitom 7:161-168

76. Heinonen A, McKay H, Whittall K, Forster B, Khan K (2001) Muscle cross-sectional area is associated with specific site of bone in prepubertal girls: a quantitative magnetic resonance imaging study. Bone 29:388-392

77. Duncan CS, Blimkie CJ, Kemp A et al (2002) Mid-femur geometry and biomechanical properties in 15-to 18-yr-old female athletes. Med Sci Sports Exerc 34:673-681

78. Adiotomre E, Summers L, Allison A et al (2017) Diagnostic accuracy of DXA compared to conventional spine radiographs for the detection of vertebral fractures in children. Eur Radiol 27:2188-2199. doi:10.1007/s00330-016-4556-3

79. Lewiecki EM, Laster AJ (2006) Clinical review: Clinical applications of vertebral fracture assessment by dual-energy x-ray absorptiometry. J Clin Endocrinol Metab 91:4215-4222. doi:10.1210/jc.2006-1178

80. Siminoski K, Lentle B, Matzinger MA, Shenouda N, Ward LM, Canadian SC (2014) Observer agreement in pediatric semiquantitative vertebral fracture diagnosis. Pediatr Radiol 44:457-466. doi:10.1007/s00247-013-2837-4

81. Alqahtani F, Messina F, Kruger E et al (2017) Evaluation of a semi-automated software program for the identification of vertebral fractures in children. Clin Radiol 72:904.e911-904.e920

82. Link TM, Guglielmi G, van Kuijk C, Adams JE (2005) Radiologic assessment of osteoporotic vertebral fractures: diagnostic and prognostic implications. Eur Radiol 15:1521-1532. doi:10.1007/s00330-005-2773-2

83. Cummings SR, Bates D, Black DM (2002) Clinical use of bone densitometry - Scientific review. Jama- J Amer Med Assoc 288:1889-1897. doi:10.1001/jama.288.15.1889

84. Damilakis J, Adams JE, Guglielmi G, Link TM (2010) Radiation exposure in X-ray-based imaging techniques used in osteoporosis. Eur Radiol 20:2707-2714. doi:10.1007/s00330-010-1845-0

85. Vokes T, Bachman D, Baim S et al (2006) Vertebral fracture assessment: The 2005 ISCD official positions. J Clin Densitom 9:37-46 doi:10.1016/j.jocd.2006.05.006

86. Hospers IC, van der Laan JG, Zeebregts CJ et al (2009) Vertebral Fracture Assessment in Supine Position: Comparison by Using Conventional Semiquantitative Radiography and Visual Radiography. Radiology 251:822-828. doi:10.1148/radiol.2513080887

87. Binkley N, Krueger D, Gangnon R, Genant HK, Drezner MK (2005) Lateral vertebral assessment: a valuable technique to detect clinically significant vertebral fractures. Osteoporos Int 16:1513-1518 doi:10.1007/s00198-005-1891-7

88. Fuerst T, Wu C, Genant HK et al (2009) Evaluation of vertebral fracture assessment by dual X-ray absorptiometry in a multicenter setting. Osteoporos Int 20:1199-1205. doi:10.1007/s00198-008-0806-9

89. Buehring B, Krueger D, Checovich M, Gemar D, Vallarta-Ast N, Genant HK, Binkley N (2010) Vertebral fracture assessment: impact of instrument and reader. Osteoporos Int 21:487-494. doi:10.1007/s00198-009-0972-4

90. Vokes TJ, Dixon LB, Favus MJ (2003) Clinical utility of dual-energy vertebral assessment (DVA). Osteoporos Int 14:871-878

91. Schousboe JT, DeBold CR (2006) Reliability and accuracy of vertebral fracture assessment with densitometry compared to radiography in clinical practice. Osteoporos Int 17:281-289

92. Rea J, Chen M, Li J et al (1999) Morphometric X-ray absorptiometry and morphometric radiography of the spine: a comparison of analysis precision in normal and osteoporotic subjects. Osteoporos Int 9:536-544

93. Ferrar L, Jiang G, Eastell R, Peel N (2003) Visual Identification of Vertebral Fractures in Osteoporosis Using Morphometric X-ray Absorptiometry. J Bone Miner Res 18:933-938

94. Crabtree NJ, Kent K (2016) Acquisition of DXA in children and adolescents. In: Bone Health Assessment in Pediatrics. Springer, pp 89-114

95. Charpak, G., (1981) La detection des particules. Recherche, *12*:1384-1396

96. Wade R, Yang H, McKenna C, Faria R, Gummerson N, Woolacott N (2013) A systematic review of the clinical effectiveness of EOS 2D/3D X-ray imaging system. Eur Spine J 22:296-304

97. Melhem, E., Assi, A., El Rachkidi, R. and Ghanem, I., 2016. EOS® biplanar X-ray imaging: concept, developments, benefits, and limitations. J Child Orthop, 10:1-14

98. Rehm J, Germann T, Akbar M, Pepke W, Kauczor H-U, Weber M-A, Spira D (2017) 3D-modeling of the spine using EOS imaging system: Inter-reader reproducibility and reliability. PLoS One 12:e0171258

99. Briot K, Fechtenbaum J, Etcheto A, Kolta S, Feydy A, Roux C (2015) Diagnosis of vertebral fractures using a low-dose biplanar imaging system. Osteoporos Int 26:2649-2655

100. Damilakis J, Maris TG, Karantanas AH (2007) An update on the assessment of osteoporosis using radiologic techniques. Eur Radiol 17:1591-1602. doi:10.1007/s00330-006-0511-z

101. Melton III LJ, Kan SH, Frye MA, Wahner HW, O'fallon WM, Riggs BL (1989) Epidemiology of vertebral fractures in women. Am J Epidemiol 129:1000-1011

102. Eastell R, Cedel SL, Wahner HW, Riggs BL, Melton LJ (1991) Classification of vertebral fractures. J Bone Miner Res 6:207-215

103. Minne H, Leidig G, Wuster C (1988) A newly developed spine deformity index (SDI) to quantitate vertebral crush fractures in patients with osteoporosis. Maturitas 10:248

104. McCloskey E, Spector T, Eyres K, Fern E, O'rourke N, Vasikaran S, Kanis J (1993) The assessment of vertebral deformity: a method for use in population studies and clinical trials. Osteoporos Int 3:138-147

105. Smith RW, Eyler WR, Mellinger RC (1960) On the incidence of senile osteoporosis. Ann Intern Med 52:773-781

106. Kleerekoper M, Parfitt A, Ellis B (1984) Measurement of vertebral fracture rates in osteoporosis. In: Copenhagen international symposium on osteoporosis, pp 103-109

107. Grados F, Fechtenbaum J, Flipon E, Kolta S, Roux C, Fardellone P (2009) Radiographic methods for evaluating osteoporotic vertebral fractures. Joint Bone Spine 76:241-247. doi:10.1016/j.jbspin.2008.07.017

108. Jensen GF, McNair P, Boesen J, Hegedüs V (1984) Validity in diagnosing osteoporosis. Observer variation in interpreting spinal radiographs. Eur J Radiol 4:1-3

109. Genant HK, Jergas M, Palermo L, Nevitt M, Valentin RS, Black D, Cummings SR (1996) Comparison of semiquantitative visual and quantitative morphometric assessment of prevalent and incident vertebral fractures in osteoporosis. J Bone Miner Res 11:984-996. doi:10.1002/jbmr.5650110716

110. Panda A, Das CJ, Baruah U (2014) Imaging of vertebral fractures. Indian J Endocrinol Metab 18:295

111. Ferrar L, Jiang G, Adams J, Eastell R (2005) Identification of vertebral fractures: An update. Osteoporos Int 16:717-728. doi:10.1007/s00198-005-1880-x

112. Koerber F, Uphoff US, Koerber S, Schönau E, Maintz D, Semler O (2012) Introduction of a new standardized assessment score of spine morphology in osteogenesis imperfecta. Fortschr Röntgenstr 184:719-725. doi: 10.1055/s-0032-1312833

113. Kim Y, Demissie S, Eisenberg R, Samelson E, Kiel D, Bouxsein M (2011) Intra- and inter-reader reliability of semi-automated quantitative morphometry measurements and vertebral fracture assessment using lateral scout views from computed tomography. Osteoporos Int 22:2677-2688

114. De Bruijne, M., Lund, M.T., Tankó, L.B., Pettersen, P.C. and Nielsen, M., (2007) Quantitative vertebral morphometry using neighbor-conditional shape models. Med Image Anal, 11:503-512

115. Cootes, T.F., Edwards, G.J. and Taylor, C.J., (2001) Active appearance models. IEEE Trans Pattern Anal Mach Intell, 6:681-685.

116. Roberts M.G., Cootes T.F., Adams J.E. (2012) Automatic Location of Vertebrae on DXA Images Using Random Forest Regression. In: Ayache N., Delingette H., Golland P., Mori K. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012. MICCAI 2012. Lecture Notes in Computer Science, vol 7512. Springer, Berlin, Heidelberg

117. Al Arif, S.M.R., Knapp, K. and Slabaugh, G., 2018, September. SPNet: Shape Prediction Using a Fully Convolutional Neural Network. In International Conference on Medical Image Computing and Computer-Assisted Intervention. pp:430-439. Springer, Cham.

118. Duman, E. and Erdem, O.A., 2018, May. A statistical edge detection framework for noisy images. In 2018 26th Signal Processing and Communications Applications Conference (SIU) . IEEE. pp: 1-4

119. Mumford, D. and Shah, J., (1989) Optimal approximations by piecewise smooth functions and associated variational problems. Commun Pure Appl Math, 42:577-685

120. Kass, M., Witkin, A. & Terzopoulos, D. (1988) Snakes: Active contour models .Int J Comput Vision 1: 321. https://doi.org/10.1007/BF00133570

121. Birch C, Knapp K, Hopkins S, Gallimore S, Rock B (2015) SpineAnalyzer (TM) is an accurate and precise method of vertebral fracture detection and classification on dual-energy lateral vertebral assessment scans. Radiography 21:278-281. doi:10.1016/j.radi.2015.02.003

Chapter Three

Alqahtani FF, Messina F, Kruger E, et al (2017) Evaluation of a semi-automated software program for the identification of vertebral fractures in children. *Clin Radiol* 72:904.e11-904.e20. https://doi.org/10.1016/j.crad.2017.04.010

3. Evaluation of a semi-automated software programme for the identification of vertebral fractures in children

Fawaz F Alqahtani[1,2], Fabrizio Messina[3], Elzene Kruger[4], Heerunpal Gill[5], Michael Ellis[5], Isla Lang[4], Penny Broadley[4], Amaka C Offiah[1,4]

Correspondence: Mr Fawaz F. Alqahtani

Room C7, C Floor, Academic Unit of Child Health
Department of Oncology and Metabolism
University of Sheffield, Medical School
Stephenson Wing
Sheffield Children's Hospital
Western Bank
Sheffield S10 2TH
Tel: 0114 271 7228
E-mail: ffmalqahtani1@sheffield.ac.uk

[1]   Academic Unit of Child Health, University of Sheffield, Sheffield, UK

[2]   Department of Radiological Sciences, College of Applied Medical Sciences, Najran University, Najran, Kingdom of Saudi Arabia

[3]   School of Health and Related Research, University of Sheffield, Sheffield, UK

[4]   Radiology Department, Sheffield Children's NHS Foundation Trust, Sheffield, UK

[5]   Sheffield Medical School, University of Sheffield, Sheffield, UK

3.1 Abstract

Purpose

There is significant inter and intraobserver variability in diagnosing vertebral fractures in children. We aimed to assess observer reliability and diagnostic accuracy in children, of a semi-automated 6-point technique developed for vertebral fracture diagnosis in adults, which records percentage loss of vertebral body height.

Methods

Using a semi-automated software programme, 5 observers independently assessed T4 to L4 from the lateral spine radiographs of 137 children and adolescents for vertebral fractures. A previous consensus read by 3 paediatric radiologists using a simplified ABQ technique (i.e. no software involved) served as the reference standard.

Results

Of a total of 1781 vertebrae, 1187 (67%) were adequately visualised by 3 or more observers. Interobserver agreement in vertebral readability for each vertebral level for five observers ranged from 0.05 to 0.47 (95% CI, -0.19, 0.76). Intraobserver agreement using the intraclass correlation coefficient (ICC) ranged from 0.25 to 0.61. Overall sensitivity and specificity were 18% (95% CI, 14 – 22) and 97% (95% CI, 97 – 98) respectively.

Conclusion

In contrast to adults, the six-point technique assessing anterior, middle and posterior vertebral height ratios is neither satisfactorily reliable nor sensitive for vertebral fractures diagnosis in children. Training of the software on paediatric images is required, in order that a paediatric standard is developed which incorporates not only specific vertebral body height ratios but also the age-related physiological changes in vertebral shape that occur throughout childhood.

3.2 Keywords

Observer agreement; Diagnostic accuracy; Osteoporosis; Paediatric; Vertebral fracture; Vertebral height

3.3 Highlights

- SpineAnalyzer™ (*Optasia Medical, Cheadle, UK,* version number: 4.0.2.19) has low observer agreement when used independently by five observers to diagnose vertebral fractures in children

- The six-point approach based on the Genant classification used by the software is not sufficiently sensitive for vertebral fractures diagnosis in children

- It may be that the placing of more than six points is required to accurately represent vertebral morphometry in children

- Development of specific paediatric software and normative values (incorporating age-related physiological variation in children) is required

3.4 Introduction

Fractures are common in childhood and repeated fractures reflect the interacting effects of low bone mineral density (BMD) and/or physical activity [1]. Vertebral fractures are a relatively common type of osteoporotic fracture. The detection of one or more vertebral compression (crush) fractures (identified by a 20% reduction in vertebral body height) is indicative of bone fragility irrespective of the reported BMD [1]. Although a lot of recent research has been conducted regarding the occurrence of osteoporotic vertebral fracture in adults, relatively less attention has been paid towards paediatric vertebral fracture, largely on account of the lack of an accepted standardised diagnostic technique in children [2].

In the absence of major trauma, reduced BMD in children and adolescents is the major cause of vertebral fracture; indeed the finding of a vertebral fracture is a main diagnostic feature of low BMD in children [1]. The low BMD may be primary (e.g. osteogenesis imperfecta) or secondary [1, 3]. For example, the STOPP studies have implicated glucocorticoids as a significant cause of secondary fractures in children and shown an incidence of vertebral fractures in those with a new diagnosis of acute lymphoblastic leukemia of 16% [4, 5]. Unlike osteoporotic fractures of the limbs, vertebral fractures are typically silent and if untreated may lead to progressive loss of vertebral body height and potential spinal deformity. If vertebral fractures are diagnosed early, however, bisphosphonate treatment can help to treat existing fractures and reduce future fracture risk [6].

Assessment of vertebral fractures in children is performed using standard lateral spine radiographs and, currently, these are interpreted using a subjective visual assessment method to identify loss of height/change in shape consistent with vertebral fracture. This approach is hampered by significant inter and intraobserver variability [2; 7, 8],

which is likely to be reduced if a more objective assessment method is applied. Semi-automated software programmes such as SpineAnalyzer™ (*Optasia Medical, Cheadle, UK)* may be the solution; but, so far, limited studies have been carried out to evaluate these programmes in children. The potential added value of these programmes is that non-radiologists may be trained to use them, freeing up radiologists' time for more specialised tasks.

The purpose of this study was to assess the observer reliability and diagnostic accuracy in children and adolescents, of the semi-automated 6-point technique developed for vertebral fracture diagnosis in adults, using a semi-automated software programme (SpineAnalyzer™, version number: 4.0.2.19*).* This software records percentage loss of vertebral body height and classifies fractures based on the Genant system [9].

3.5 Materials and methods

3.5.1 Study population

This study involved the retrospective analysis of images obtained as part of a larger prospective study of 250 children recruited between November 2011 and February 2014 [7]. All images used in this study were of patients recruited from our single center. The mean age of the 137 subjects at the time of image acquisition was 12.0 years (range 5 to 15) and 45 (33%) were male. The majority, 199 (80%) had suspected reduction in BMD (including children with osteogenesis imperfecta, inflammatory bowel disease, rheumatologic conditions, cystic fibrosis and celiac disease). The remaining 51 (20%) patients were recruited from spine clinic.

3.5.2 Retrospective power calculations

A power calculation was done comparing the sensitivity and specificity with those likely to have occurred by random using kappasize package R gave us that a sample size greater than 119 would be required with significance=0.05 and power at 80%. The sample size of this study (137) is greater than this [10].

3.5.3 Lateral spine imaging

Lateral images of the thoracolumbar spine were acquired using one of two Phillips Healthcare machines *(TH3 Digital or TH Bucky Diagnost, Guildford, UK)* following the European guidelines for imaging the spine in children as previously described [7]. The subjects were asked to remain in the lateral decubitus position with flexed knees and hips. Depending on the size of each child being examined, thoracolumbar or separate thoracic and lumber spine images were obtained. As outlined in a previous study, the tube-to-film distance was set at 100 cm, and the films were centered at T7 and L3 for the thoracic and lumbar views, respectively [11]. The average exposures for thoracic, lumbar and thoracolumbar spine radiographs were 75, 84 and 74kV respectively.

3.5.4 Image analysis

Lateral spine images were analysed independently by five observers (a radiologist, two radiographers, and two medical students), who attempted readings for all 137 cases, with each observer being blinded to the other evaluations. Prior to commencing the study, the four non-radiologists were trained on use of the software by the radiologist, learning from non-study spine radiographs. All observers' technique were observed and assessed in person by the radiologist and feedback provided to them orally. A previous consensus arrived at by three paediatric radiologists using a simplified

algorithm-based qualitative (ABQ) technique (i.e. with no software involved) served as the reference standard [11].

As the first step in the semi-automated analysis using SpineAnalyzer™ (version number: 4.0.2.19), observers identify T4 to L4 vertebral bodies by placing a point at or close to the center of each vertebral body and indicating to the software the highest identified vertebral body (for example, T4). Having indicated T4, the software programme recognises all identified vertebral bodies between T4 and L4 and automatically identifies six points corresponding to the four corners and the midpoints of the superior and inferior endplates of each vertebral body – observers modify the placement of these points as necessary. The software does not recognise vertebral bodies above T4 or below L4 (Figure 3.1).

Image
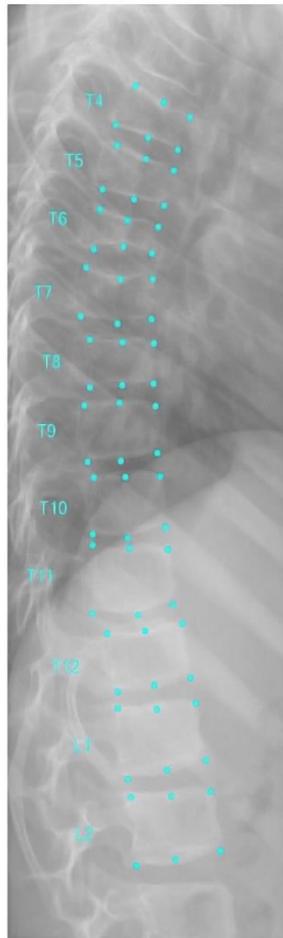
Deformity Results

Level	Morphometry (% Deformation)			Grade	SQ Score
	Wedge	Bicon.	Crush		
T4	20.6%	11.5%	3.3%	Mild	-
T5	22.6%	10.8%	-6.0%	Mild	-
T6	32.6%	35.4%	17.5%	Moderate	-
T7	18.3%	9.6%	-3.2%	Normal	-
T8	16.2%	9.7%	14.8%	Normal	-
T9	15.6%	-3.4%	0.4%	Normal	-
T10	8.9%	-7.6%	6.3%	Normal	-
T11	19.5%	2.3%	-15.7%	Normal	-
T12	6.2%	5.4%	5.2%	Normal	-
L1	18.4%	13.0%	-3.3%	Normal	-
L2	13.0%	7.6%	-3.1%	Normal	-
L3	-	-	- -		-
L4	-	-	- -		-

0 - Normal < 20%	1 - Mild 20% < 25%	2 - Moderate 25% < 40%	3 - Severe ≥ 40%

Level	Differential Diagnosis
T4	-
T5	-
T6	-
T7	-
T8	-
T9	-
T10	-
T11	-
T12	-
L1	-
L2	-
L3	-

Figure 3.1 Lateral thoracolumbar spine radiograph, illustrating the six semi-automatically identified points used to outline the vertebral bodies and the deformity result produced by the SpineAnalyzer™ programme

Following placement of the six points, anterior, middle and posterior vertebral heights are automatically determined by the software. With the help of these measurements, the anterior: posterior, middle: posterior, posterior: posterior[+1] and posterior: posterior [-1] height ratios are calculated (+1 and -1 indicate the vertebrae immediately above (+1) and below (-1) the vertebra of interest). The vertebral bodies are then classified according to their height ratios, based on the scoring system developed by Genant (Table 3.1 and Figure 3.1) [9].

Table 3.1 Genant grading system for vertebral fracture (VF) [9]

Loss of Height (%)	Grade	Interpretation
≤ 20%	0	No VF
21 to 25%	1	Mild VF
26 to 40%	2	Moderate VF
≥ 41%	3	Severe VF

For the purposes of this study, since the assessment only included lateral spine images, to maintain the consistency of vertebral level assignment between the five observers, the first vertebral body not associated with ribs was labelled as L1, while the lowermost vertebral body associated with ribs was labelled as T12. All subjects had 5 lumber vertebrae [12] and if the observer was unable to identify T12 and/or L1, (e.g. due to parallax distortion), then that image was not scored. The strengths and weaknesses of the Genant semi quantitative (SQ) and the algorithm based qualitative (ABQ) methods are summarised in Table 3.2.

Table 3.2 Strengths and weaknesses of the Genant semi quantitative (SQ) and the algorithm based qualitative (ABQ) methods

Scoring system	Strengths	Weakness

Genant SQ	- Simple and quick
	- Widely used in clinical trials
	- Approved and used by international institutions such as International Osteoporosis Foundation (IOF) and ISCD
		- Subjective
		- Vertebral deformity has to be greater than 20%
		- No clear description of differential diagnosis of deformity:
		  • Osteoporosis vertebral fracture
		  • Traumatic vertebral fracture
		  • Non- fracture deformity

ABQ	- Requires evidence of fracture of the vertebral endplate - No minimum threshold for apparent reduction in vertebral height - Allowance is made for variation in vertebral dimensions at different vertebral levels or short vertebral heights e.g. Scheuermann's disease, scoliosis etc...	- Training, skilled reader and experience needed

3.5.5 Statistical analysis

R software was employed for data analysis [13]. The frequency of readable vertebrae for each observer and for all vertebrae from T4 to L4 was calculated.

Diagnostic accuracy (sensitivity, specificity and 95% confidence interval) calculations of the observers' readings were calculated by comparing with a previously established consensus arrived at by three experienced paediatric radiologists using a simplified ABQ scoring system, Table 3.3 [11]. For diagnostic accuracy calculations, both sABQ

and SpineAnalyzer™ scores of 0 and 1 were interpreted as, "no clinically significant fracture". Inter and intra observer agreement were calculated using kappa and intraclass correlation coefficient (ICC) respectively [14, 15].

Table 3.3 Simplified algorithm based qualitative scoring system [11]

Parameter	Score	Interpretation
Height	0	Normal height
	1	Height loss ≤ 24%
	2	Height loss ≥ 25%
Endplates	a	Normal endplates
	b	Single endplate affected
	c	Both endplates affected
Other	3	Non-osteoporotic (non-fracture) deformity
	4	Uncertain or unable to determine due to image quality

3.5.6 Approvals

Local Research Ethics Committee approval was obtained for the main study from which the images were drawn but was not separately required for this study. The study was registered with our Research and Innovation Department prior to commencement.

3.6 Results

3.6.1 Prevalence of fractures

Overall, 20 (15 %) patients had one or more vertebral fracture (vertebral height loss 20 % or more). Per-vertebra, 48 vertebral fractures were identified by three or more observers using SpineAnalyzer™. The majority of these fractures were in the mid-thoracic region, with T7 being the most fractured level - 9 (19%).

3.6.2 Readability of radiographic lateral spine images within SpineAnalyzer™ software programme

Of the possible total 1781 vertebrae, from T4 through to L4 (i.e. 13 vertebrae per subject in 137 subjects), 1310 (73.55%) were adequately visualised by Observer 1, 1370 (77%) by Observer 2, 1376 (77%) by Observer 3 and 1319 (74%) and 1344 (75%)

by Observers 4 and 5 respectively (Figure 3.2). A total of 1187 (67 %) were adequately visualised by three or more observers, permitting comparison of morphology results. The visibility was relatively limited in the upper part of the thoracic spine; T4 was the least readable level, being adequately visualised by all observers on 423 (62%) radiographs.
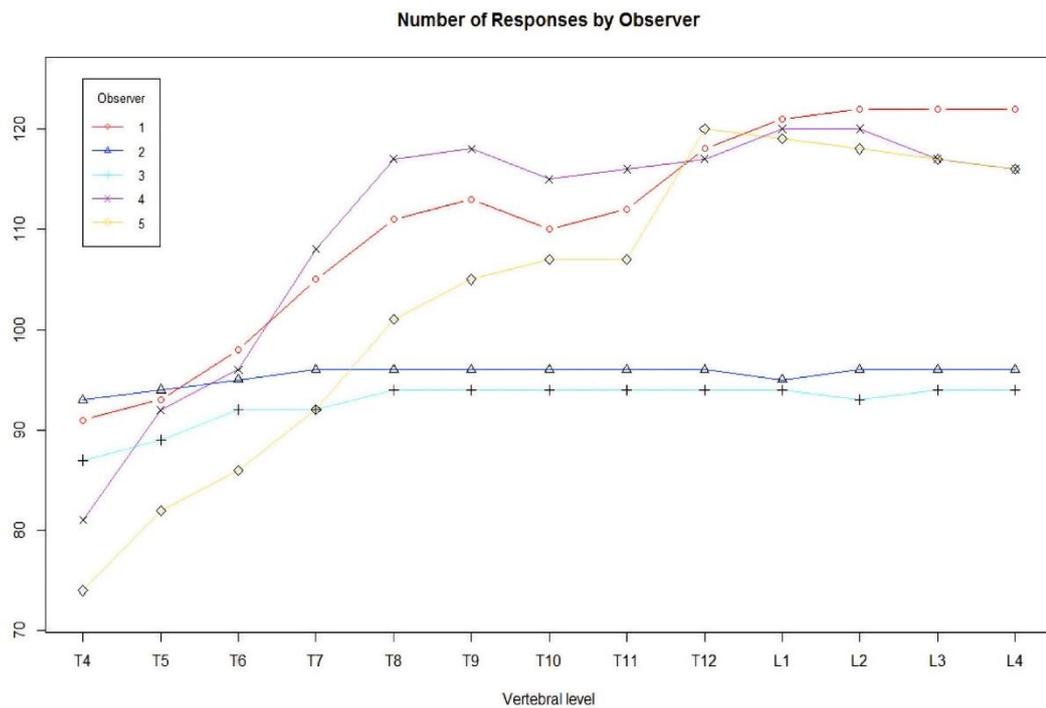
Figure 3.2 Number of readable vertebrae for each observer.   There is a trend towards increasing readability from the upper thoracic to the lumbar spine

Sensitivity and specificity values of the observers' readings with their 95% confidence intervals are presented in Table 3.4. T6 had the highest and L3 the lowest sensitivity, while L4 had the highest and T11 the lowest specificity. Overall sensitivity was 18% (95% CI, 14 – 22), while overall specificity was 97% (95% CI, 97 – 98).

Table 3.4 Sensitivity, specificity, interobserver (kappa) and intraobserver (ICC) reliability of SpineAnalyzer™ for vertebral fracture diagnosis in children

	Number of truth fractures/Number of readable vertebrae	Sensitivity	95% CI	Specificity	95% CI	% Agreement	Kappa Score (k)	95% CI	ICC	SD
T4	25/83	18.20%	13.22 -23.18	96.27%	90.91 - 100.00	63.24	0.05	-0.19 - 0.30	0.30	0.43
T5	32/87	20.56%	11.25 -29.87	98.75%	96.33 - 100.00	89.32	0.36	0.13 - 0.59	0.61	0.01
T6	31/94	32.20%	16.70 - 47.70	96.51%	91.29 - 100.00	86.53	0.39	0.12 - 0.65	0.43	0.07
T7	35/94	17.25%	7.25 - 27.25	96.06%	92.51 - 99.60	88.22	0.24	-0.03 - 0.51	0.56	0.04
T8	29/104	22.80%	12.21 - 33.3	95.70%	92.56 - 98.84	89.46	0.41	0.11 - 0.72	0.54	0.02
T9	20/105	18.22%	6.39 - 30.04	98.25%	96.90 - 99.61	92.85	0.23	-0.07 - 0.53	0.38	0.05
T10	16/103	10.17%	3.40 - 16.94	98.10%	96.43 - 99.77	95.00	0.17	-0.12 - 0.46	0.33	0.12
T11	18/106	19.50%	9.11 - 29.89	95.62%	92.40 - 98.84	91.35	0.23	-0.05 - 0.52	0.25	0.07
T12	17/110	24.74%	13.23 - 36.24	96.04%	92.98 - 99.10	92.58	0.47	0.18 - 0.76	0.48	0.18
L1	12/110	17.30%	9.97 - 24.67	98.32%	95.88 - 100.00	94.21	0.34	0.07 - 0.62	0.52	0.07
L2	16/112	13.70%	3.61 - 23.78	98.59%	96.57 - 100.00	95.49	0.34	-0.02 - 0.70	0.54	0.06
L3	19/111	7.93%	2.41- 13.45	99.06%	97.87 - 100.00	97.53	0.39	-0.02 - 0.80	0.46	0.01
L4	15/111	10.30%	3.41 - 17.20	99.32%	98.56 - 100.00	97.03	0.26	-0.15 - 0.67	0.43	0.06

Note: T4, T5, T6, L1, L2, L3 and L4 95% CI upper bound were capped at 100.00

The average kappa for interobserver agreement in respect to vertebral readability between the five observers for each of the 13 vertebrae ranged from 0.05 to 0.47 (95% CI, -0.19, 0.76). Table 3.3 shows the agreement (average kappa score) between the five observers using SpineAnalyzer™. T4 had the lowest and T12 the highest agreement. Average intraobserver agreement ranged from 0.25 to 0.61. Table 3.4 also shows that overall, there was poor/fair agreement for the 13 vertebrae, with the only exception being T5, for which agreement was good. Table 3.5 compares results of this current study with those of the only other study to date that has assessed the 6-point technique in children [8] and with those of the largest published study to compare VFA with radiographs for diagnosis of vertebral fracture in children [7].

Table 3.5 Summary of diagnostic accuracy and observer reliability of SpineAnalyzer™ in children

Modality/Method	Sensitivity		Specificity		Interobserver Reliability		Intraobserver Reliability	
	Adiotomre [7]	Crabtree [8]	Adiotomre [7]	Crabtree [8]	Adiotomre [7]	Crabtree [8]	Adiotomre [7]	Crabtree [8]
Radiographs	74%-84%	_	72%-96%	_	0.39-0.46	0.53-0.59	0.51-0.70	_
DXA VFA	70%-78%	66%-81%	72 %-97%	78%-99%	0.32-0.50	0.60-0.66	0.43-0.77	_
6-point analysis*	14%-22%	62%-78%	97%-98%	70%-99%	0.05-0.47	0.36-0.41	0.25-0.61	_

* Current study: Analysed radiographs, Crabtree et al: Analysed DXA

Figure 3.3 illustrates examples of good and poor observer agreement, while Figure 3.4 illustrates differences in diagnostic outcome due to early ossification of the apophyses causing minor observer differences in placement of the six points. Figure 3.5 demonstrates false positive and false negative results of SpineAnalyzer™
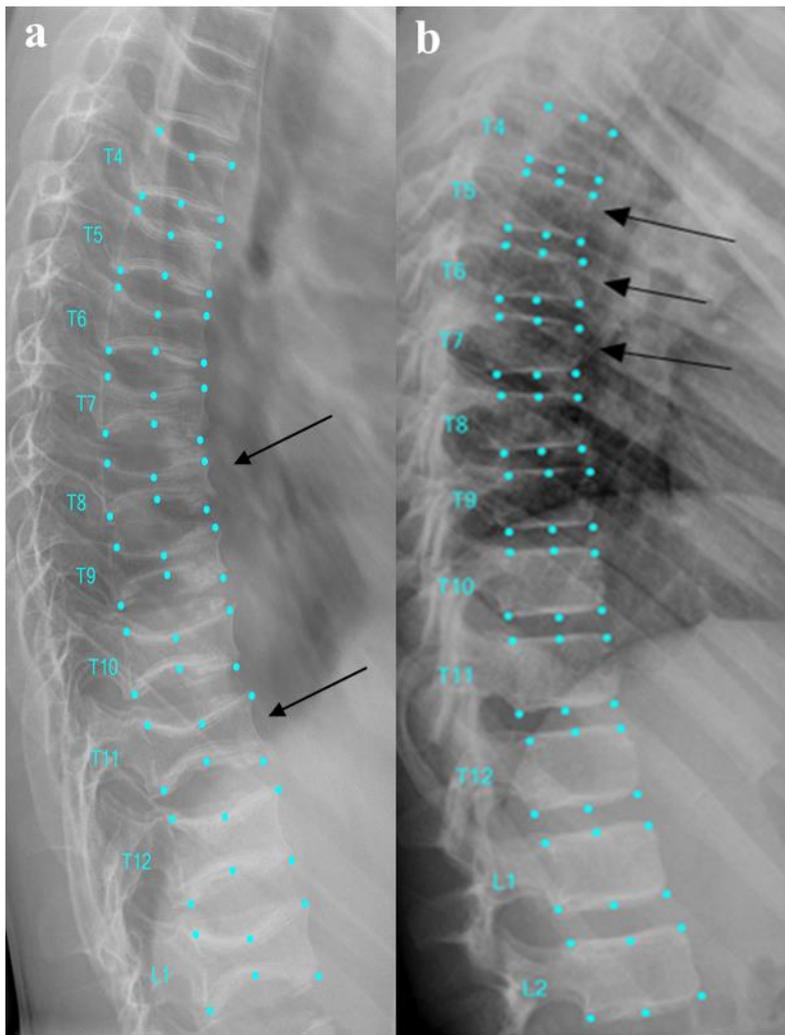
Figure 3.3a Observer agreement: all five observers identified a severe T8 fracture. Similarly, the T11 fracture was identified by all, but graded as mild by two observers, moderate by one and severe by two

Figure 3.3 b Lack of observer agreement: T5 - T7 were deemed non-evaluable by one observer and graded as no fractures by one observer, mild fractures by two and moderate fractures by one
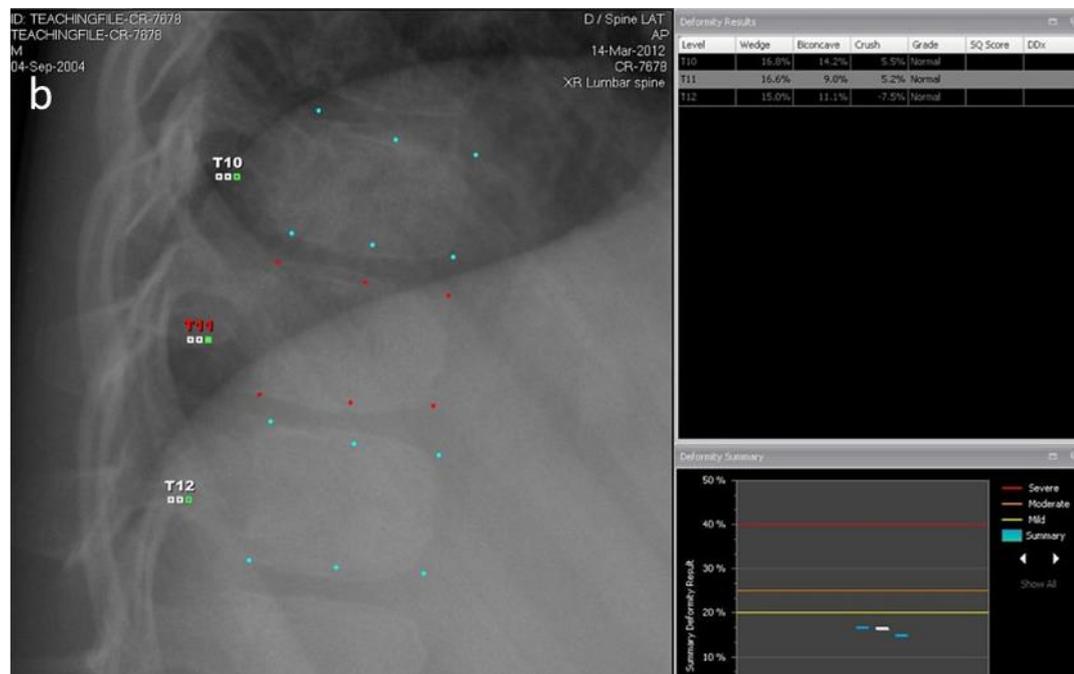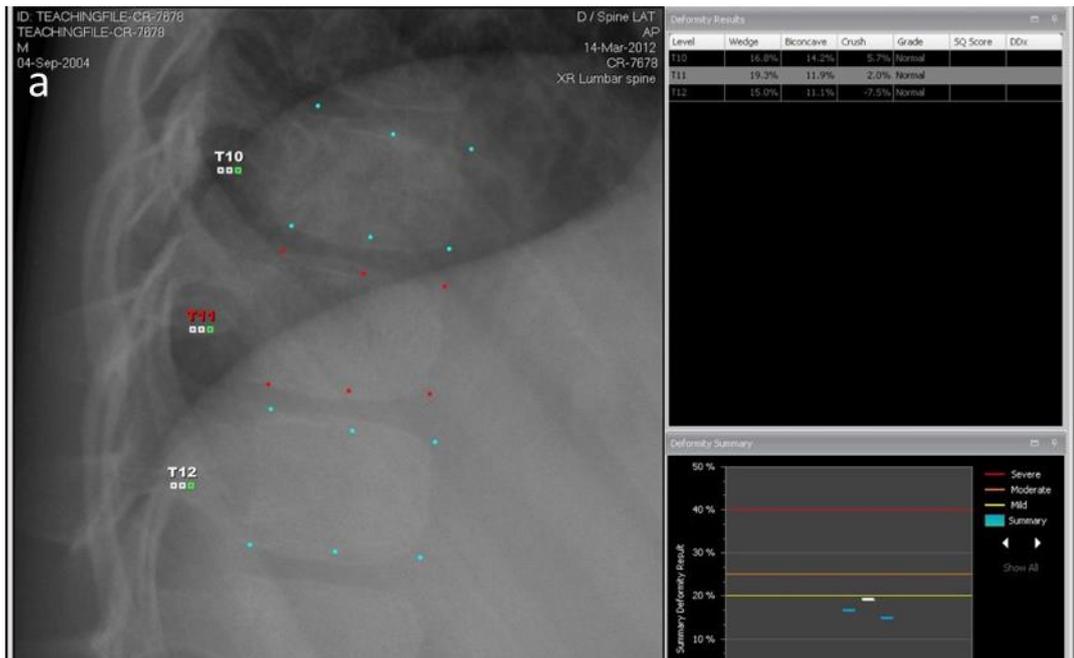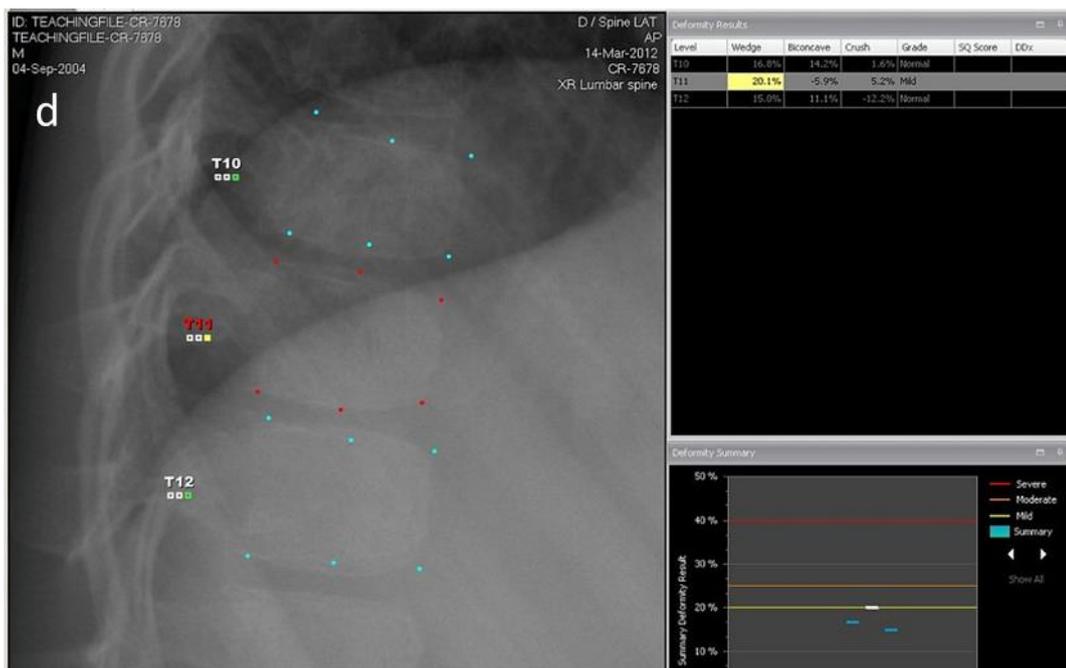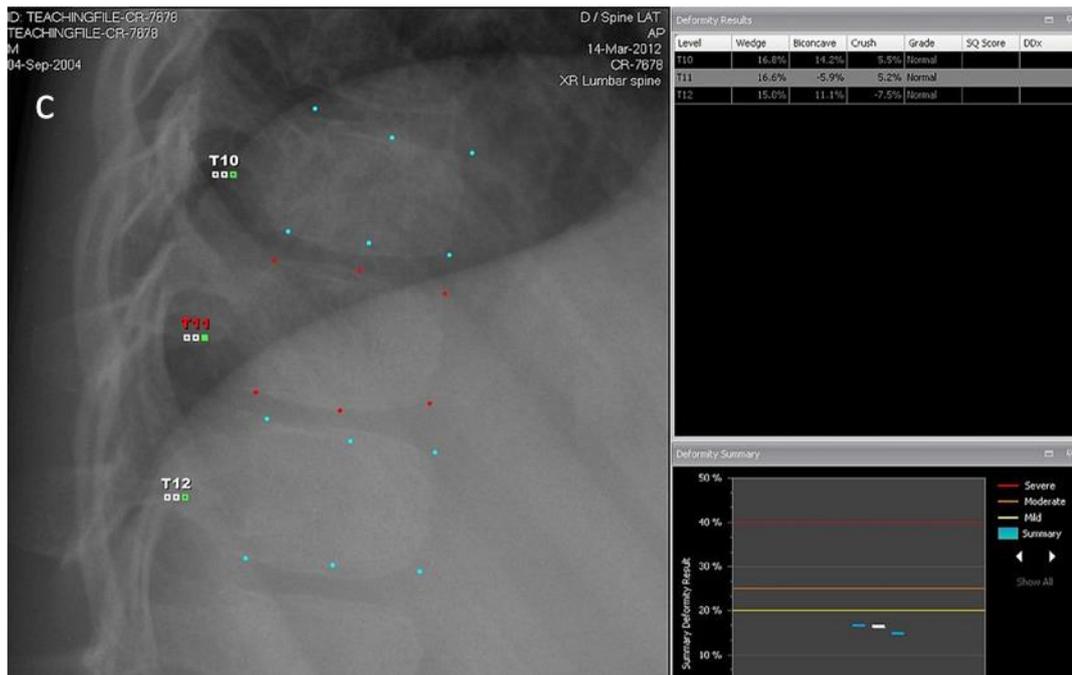
Figure 3.4 Effect of minor alterations in point placement for T11 in the same patient in which there is early apophyseal ossification. a (no manipulation), b (posterior manipulation) and c (middle manipulation) were classified by SpineAnalyzer™ as normal, while d (anterior manipulation) was scored by SpineAnalyzer™ as a mild fracture

Image

Deformity Results

Level	Morphometry (% Deformation)			Grade	SQ Score
	Wedge	Bicon.	Crush		
T4	6.1%	-0.2%	-2.3%	Normal	-
T5	1.4%	3.3%	-2.6%	Normal	-
T6	8.6%	4.6%	-4.1%	Normal	-
T7	27.6%	12.6%	-0.2%	Moderate	-
T8	21.7%	10.5%	2.4%	Mild	-
T9	17.0%	9.6%	4.6%	Normal	-
T10	10.4%	7.9%	-10.6%	Normal	-
T11	-1.8%	-5.5%	-2.7%	Normal	-
T12	7.1%	8.3%	16.6%	Normal	-
L1	7.7%	9.5%	8.0%	Normal	-
L2	6.5%	2.1%	-10.1%	Normal	-
L3	-	-	-	-	-
L4	-	-	-	-	-

0 - Normal < 20%	1 - Mild 20% < 25%	2 - Moderate 25% < 40%	3 - Severe ≥ 40%

Level	Differential Diagnosis
T4	-
T5	-
T6	-
T7	-
T8	-
T9	-
T10	-
T11	-
T12	-
L1	-
L2	-
L3	-

Figure 3.5a False positive SpineAnalyzer™ result. Wedging of T7 and T8 as indicated by SpineAnalyzer™ was reported by the consensus expert panel as physiological, rather than pathological wedging

Image
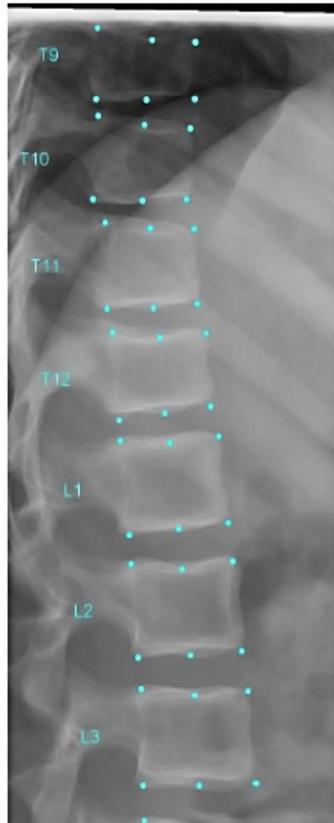
Deformity Results

| Level | Morphometry (% Deformation) | | | | SQ Score |
	Wedge	Bicon.	Crush	Grade	
T4	-	-	-	-	-
T5	-	-	-	-	-
T6	-	-	-	-	-
T7	-	-	-	-	-
T8	-	-	-	-	-
T9	19.9%	14.5%	16.0%	Normal	-
T10	15.7%	9.6%	2.2%	Normal	-
T11	13.7%	7.4%	-1.0%	Normal	-
T12	16.7%	13.4%	7.8%	Normal	-
L1	8.9%	8.9%	-2.4%	Normal	-
L2	2.9%	7.2%	0.0%	Normal	-
L3	4.6%	7.0%	-1.8%	Normal	-
L4	1.3%	8.1%	0.0%	Normal	-

0 - Normal < 20%	1 - Mild 20% < 25%	2 - Moderate 25% < 40%	3 - Severe ≥ 40%

Level	Differential Diagnosis
T4	-
T5	-
T6	-
T7	-
T8	-
T9	-
T10	-

Figure 3.5b False negative SpineAnalyzer™ result. T11, T12 and L2 were reported by the consensus expert panel as fractured but were scored normal by SpineAnalyzer™

3.7 Discussion

One or multiple vertebral fracture without high-energy trauma or local disease is indicative of osteoporosis in children. Early and accurate diagnosis is important to allow appropriate treatment to commence.

There is a relatively low observer reliability for current techniques of vertebral fracture diagnosis in children; with reported kappa values for inter and intraobserver reliability ranging from 0.39 to 0.59 and 0.33 to 0.84 respectively [2,7,8]. A recent study in adults showed an agreement between SpineAnalyzer™ and readers ranging from 0.96 to 0.97 [16]. The authors suggested that SpineAnalyzer™ is an accurate tool for measuring vertebral height and identifying vertebral fractures in adults. The purpose of this current study was to evaluate the accuracy and reliability of the semi-automated 6-point technique for diagnosing vertebral fracture in children. To our knowledge, this evaluation is the largest to assess vertebral morphometry in children using semi-automated 6-point technique software, with only one other study on the same subject published to date [8].

Compared to our results, observer reliability has been shown to be higher in studies of the diagnostic accuracy of vertebral fracture detection in adults using both visually-based scoring systems and software [16-19]. A recent study on children [2], based on the observation of radiographic images utilising Genant's semi-quantitative (SQ) technique, showed higher inter-kappa agreement for vertebral fracture diagnosis (k=0·45 to 0·54) than both our corresponding SpineAnalyzer™ calculations (k = 0.05 to 0.47) and those of Crabtree et al (k = 0.36 to 0.41) [8]. Results of the three studies should be directly comparable, given that the SpineAnalyzer™ categories are based on Genant's scoring system. It seems that small differences between observers in

point placement account for the reduced observer reliability of SpineAnalyzer™, compounded by the fact that the final categorisation is based on ratios and not simple measurements. This is supported by the fact that the paediatric study from which images for this report were drawn also obtained a higher level of interobserver agreement (k = 0·394 to 0·455) when utilising a simplified ABQ technique for vertebral morphometry [11].

Agreement between the observers reached a maximum kappa of 0.47 (95% CI, 0.18, 0.76) with the greatest level of agreement being at T12 and L4 (fair to moderate) whilst the least was at T4 (slight to poor). At each vertebral level, there was diversity in the interobserver agreement and readability of the vertebra (Figure 3.4). Results suggest that the observers could visualise the lower vertebral levels for point placement more adequately and that the calculations were correspondingly more precise than those made for the upper vertebral levels, underlining the difficulty in applying SpineAnalyzer™ for the upper thoracic spine. These findings support those of previous studies reporting that identification of vertebrae in the mid and upper thoracic spine is one of the major challenges in identifying vertebral fracture in children [2; 3].

Reasons for poor visibility include the summation caused by intrathoracic tissues and shoulders; poor image quality; and patient positioning. Therefore, the patient positioning protocol and radiographic parameters selected for imaging larger patients play an important role in improving image quality and visibility, in order that upper thoracic vertebrae can be assessed. In this regard, it should be noted that lateral spine DXA allows improved visibility of the upper thoracic spine compared to radiographs [7], which may account for the improved observer reliability of SpineAnalyzer™ in the study by Crabtree et al [8] compared to this current study.

In this study, only radiographs were used, but as discussed previously in the literature review (page 47), although dual-energy and single-energy VFA modes can help in identifying vertebral fractures, they have strengths and drawbacks. For example, the Hologic has both dual-energy and single-energy VFA modes. Vertebrae were well visualised on single-energy scans, but the increasing effect of soft tissue artefacts as one moves up the chest can affect visualisation of the thoracic vertebrae [20]. The image, therefore, requires more practice and experience to analyse successfully [21]. The recent scanner provided by GE Healthcare's DVA may give us an effective VFA diagnostic tool by combining dual-energy scans with single-energy scans and thus, single-energy and dual-energy combination may help for further increase in the number of vertebrae visualised (a typical scan can be analysed up to T5). It should be noted that, although the scan time for competing single-energy mode is short, analysis may be affected by soft tissue artefacts in the image.

Finally, variability in observer reliability may be related to differences in identifying T12/L1. In future studies, this limitation can be countered by having a marker placed adjacent to an agreed vertebra so that all observers recognise the same vertebral levels. This step should be done prior to study commencement at the stage of image anonymisation. As the lead researcher, I will choose the level of T12 (the lowermost vertebral body associated with a pair of ribs was always designated as T12) for all images and place a marker (e.g. using Photoshop programme). This will then be confirmed by an experienced paediatric radiologist.

Compared to the consensus read of the radiological experts, overall sensitivity of the semi-automated 6-point technique was only 18% (95%CI of 14 – 22) while overall specificity was 97% (95%CI of 97 – 98). These findings are likely a result of a high degree of subjectivity in placing the original six semi-automated points used by the software to identify vertebral fracture. This is despite the training given prior to commencing the study. The sensitivity results may also be low because identifying vertebral fracture using SpineAnalyzer™ is based only on the loss of height of vertebral bodies, while the simplified ABQ method is a visual method which considers alterations in the vertebral endplates that may be non-fracture related. Interpretation of SpineAnalyzer™ measurements is based on a grading system derived from analysis of thoracolumbar spine radiographs of 57 postmenopausal women and developed for adults [8]. Nevertheless, the Genant scoring system has been used with satisfactory results in a number of paediatric studies [22, 23] and therefore we suggest that the placement of only 6 points is insufficient to capture vertebral morphometry in children and placement of further points may be required.

Another factor that affects sensitivity of the software is observer skill and experience. Although in theory no medical knowledge/specialised skills are required to identify the four corners of the vertebral bodies and center of inferior and superior endplates, small differences in placement affect the overall height ratios and factors confounding point placement and/or fracture categorisation include visibility of vertebrae, early ossification of apophyses, physiological wedging and non-fracture related irregularities of vertebral endplates. Observers in this study included a musculoskeletal consultant radiologist, 2 radiographers and 2 medical students. Despite the training received, the disparate experience of the observers may be a

weakness of the study, particularly given the confounding influence of physiological variations on point placement. This will need to be considered if such programmes are to be used for role extension. If the 6-point or any semi-automated systems are to be more accurate and reliable, then a precise algorithm is required describing where the points should be placed if, for example, the apophyses are unossified and having ossified, prior to fusion. The difficulty in reproducible point-placement is also reflected by the low intraobserver reliability, even for the experienced radiologist. While the purpose of this current study was specifically to address the reliability of SpineAnalyzer™ amongst non-radiologists, in retrospect, and particularly given the poor observer reliability, it would have been interesting to have recruited and compared the results of at least two paediatric (or musculoskeletal) radiologists. This limitation of the current study is a future objective.

3.8 Conclusion

We conclude that although it appears useful in adults, from whose radiographs and for whom it was developed, due to its low inter and intraobserver reliability and sensitivity, currently the six-point technique comparing vertebral height ratios is neither satisfactorily accurate nor reliable for vertebral fracture diagnosis in children. We suggest that the system needs training on paediatric images, with a specific algorithm designed to determine point placement, incorporate overall vertebral body shape and that the classification be based on a grading system specifically designed to differentiate physiological variation from vertebral fracture.

3.9 Acknowledgment

3.10 References

1. Bishop N, Arundel P, Clark E, et al (2014) Fracture Prediction and the Definition of Osteoporosis in Children and Adolescents: The ISCD 2013 Pediatric Official Positions. J Clin Densitom 17:275-280

2. Siminoski K, Lentle B, Matzinger MA, Shenouda N, Ward LM, Canadian SC (2014) Observer agreement in pediatric semiquantitative vertebral fracture diagnosis. Pediatr Radiol 44:457-466

3. Kyriakou A, Shepherd S, Mason A, Ahmed SF (2015) A critical appraisal of vertebral fracture assessment in paediatrics. Bone 81:255-259

4. Huber AM, Gaboury I, Cabral DA, et al (2010) Prevalent vertebral fractures among children initiating glucocorticoid therapy for the treatment of rheumatic disorders. Arthritis Care & Research 62:516-526. doi: 10.1002/acr.20171. PubMed PMID: 20391507; PubMed Central PMCID: PMC3958950

5. Halton J, Gaboury I, Grant R, et al (2009) Advanced vertebral fracture among newly diagnosed children with acute lymphoblastic leukemia: results of the Canadian steroid-associated osteoporosis in the pediatric population (STOPP) research program. J Bone Miner Res 24:1326-1334. Epub 2009/02/13. doi: 10.1359/jbmr.090202 10.1359/jbmr.090202 [pii]. PubMed PMID: 19210218

6. Shaw NJ. Management of osteoporosis in children (2008) Eur J Endocrinol 159:S33-S39.

7. Adiotomre E, Summers L, Allison A, et al (2016) Diagnostic accuracy of DXA compared to conventional spine radiographs for the detection of vertebral fractures in children. Eur Radiol 27:88-99

8. Crabtree N, Chapman S, Hogler W, Hodgson K, Chapman D, Bebbington N, Shaw NJ (2017) Vertebral fractures assessment in children: Evaluation of DXA imaging versus conventional spine radiography. Bone 97:168 -74 doi http://dx. doi.org/10. 1016/j.bone.2017.01.006

9. Genant HK, Wu CY, Vankuijk C, Nevitt MC (1993) Vertebral Fracture Assessment Using A Semiquantitative Technique. J Bone Miner Res 8:1137-1148

10. Buderer, N.M.F (1996) Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. Acad Emerg Med, 3:895-900

11. Adiotomre E, Summers L, Allison A, et al (2016) Diagnosis of vertebral fractures in children: is a simplified algorithm-based qualitative technique reliable? Pediatr Radiol 46:680-688

12. Peel, N.F.A., Johnson, A., Barrington, N.A., Smith, T.W.D. and Eastell, R., (1993) Impact of anomalous vertebral segmentation on measurements of bone mineral density. J Bone Miner Res 8:719-723

13. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. Accessed 25 January 2017

14. Landis JR, Koch GG (1977) Measurement Of Observer Agreement For Categorical Data. Biometrics 33:159-174

15. Cicchetti DV (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 6:284

16. Birch C, Knapp K, Hopkins S, Gallimore S, Rock B (2015) SpineAnalyzer (TM) is an accurate and precise method of vertebral fracture detection and classification on dual-energy lateral vertebral assessment scans. Radiography 21:278-281

17. Ferrar L, Jiang G, Schousboe JT, DeBold CR, Eastell R (2008) Algorithm-based qualitative and semiquantitative identification of prevalent vertebral fracture: Agreement between different readers, imaging modalities, and diagnostic approaches. J Bone Miner Res 23:417-424

18. Kim YM, Demissie S, Eisenberg R, Samelson EJ, Kiel DP, Bouxsein ML (2011) Intra-and inter-reader reliability of semi-automated quantitative morphometry measurements and vertebral fracture assessment using lateral scout views from computed tomography. Osteoporos Int 22:2677-2688

19. Oei L, Ly F, El Saddy S, et al (2013) Multi-functionality of computer-aided quantitative vertebral fracture morphometry analyses. Quant Imaging Med Surg 3:249-255

20. Rea, J.A., Li, J., Blake, G.M., Steiger, P., Genant, H.K. and Fogelman, I. (2000) Visual assessment of vertebral deformity by X-ray absorptiometry: a highly predictive method to exclude vertebral deformity. Osteoporos Int 11:660-668

21. Rea, J.A., Steiger, P., Blake, G.M. and Fogelman, I. (1998) Optimizing data acquisition and analysis of morphometric X-ray absorptiometry. Osteoporos Int 8:177-183

22. Lentle B, Ma J, Jaremko JL, et al (2016) The Radiology of Vertebral Fractures in Childhood Osteoporosis Related to Glucocorticoid Administration. J Clin Densitom. 19:81-88

23. Cummings EA, Ma J, Fernandez CV, et al (2015) Incident Vertebral Fractures in Children With Leukemia During the Four Years Following Diagnosis. J Clin Endocrinol Metab 100:3408-3417

Chapter Four

4. Are semi-automated software programmes designed for adults accurate for the identification of vertebral fractures in children?

Fawaz F Alqahtani[1,2], Fabrizio Messina[3], and Amaka C Offiah[1,4]

Correspondence: Mr Fawaz F. Alqahtani

Room C7, C Floor, Academic Unit of Child Health
Department of Oncology and Metabolism
University of Sheffield, Medical School
Stephenson Wing
Sheffield Children's Hospital
Western Bank
Sheffield S10 2TH
Tel: 0114 271 7228
E-mail: ffmalqahtani1@sheffield.ac.uk

[1] Academic Unit of Child Health, University of Sheffield, Sheffield, UK

[2] Department of Radiological Sciences, College of Applied Medical Sciences, Najran University, Najran, Kingdom of Saudi Arabia

[3] School of Health and Related Research, University of Sheffield, Sheffield, UK

[4] Radiology Department, Sheffield Children's NHS Foundation Trust, Sheffield, UK

4.1 Abstract

Purpose

To assess whether diagnostic accuracy and observer reliability of morphometric vertebral fracture analysis (MXA) in children can be improved using AVERT™, a semi-automated programme developed for vertebral fracture diagnosis in adults, which utilises a 33-point technique to record percentage loss of vertebral body height compared to a 6-point technique (SpineAnalyzer™), which has previously been shown to be of insufficient reliability.

Methods

Lateral spine radiographs (XR) and dual energy X-ray absorptiometry (DXA) scans - vertebral fracture assessment (VFA) of 50 children and young people (performed on the same day on the same children) were analysed by two observers using two different programmes (AVERT™ and SpineAnalyzer™). Diagnostic accuracy (sensitivity, specificity, false negative (FN) and false positive rates (FP)) was calculated by comparing with a previously established consensus arrived at by three experienced pediatric musculoskeletal radiologists, using a simplified algorithm based qualitative scoring system. Observer agreement was calculated using Cohen's kappa.

Results

On radiographs, overall sensitivity, specificity, FP and FN rates of AVERT™ and SpineAnalyzer™ were 36%, 95%, 5% and 64% and 26%, 98%, 2% and 75% respectively. On DXA, overall sensitivity, specificity, FP and FN rates of AVERT™ and SpineAnalyzer were 41%, 91%, 9% and 59% and 31%, 96%, 4% and 69% respectively. Reliability (Kappa) of AVERT™ and SpineAnalyzer™ ranged from

0.34 to 0.37 (95% CI, 0.26 – 0.46) and 0.26 to 0.31 (95% CI, 0.16 – 0.44) respectively. Inter and intraobserver agreement for both programmes using kappa ranged from 0.41 to 0.47 and 0.50 to 0.79 respectively.

Conclusion

The 33-point technique has slightly higher accuracy for the representation of vertebral morphometry in children when compared to the 6-point technique. However, neither AVERT™ nor SpineAnalyzer™ are satisfactorily reliable for vertebral fracture diagnosis in children.

4.2 Keywords

Observer agreement; Diagnostic accuracy; Osteoporosis; Paediatric; Vertebral fracture; Vertebral height

4.3 Highlights

- SpineAnalyzer™ and AVERT™ have low diagnostic accuracy and observer agreement when compared to three paediatric radiologists' readings for the diagnosis of vertebral fractures in children

- Neither AVERT™ nor SpineAnalyzer™ is satisfactorily reliable for vertebral fracture diagnosis in children

- Development of specific paediatric software and normative values (incorporating age-related physiological variation in children) is required

4.4 Introduction

Low bone mass is characterised by structural deterioration of bone tissue, leading to bone fragility and an increased susceptibility to fractures, especially of the spine and long bones. According to the International Society for Clinical Densitometry (ISCD) one or multiple vertebral fractures - identified by a 20% reduction in vertebral body height - indicates bone fragility, in the absence of local disease or significant trauma [1].

Osteoporotic vertebral fractures are increasingly recognised in children as a vital sign of low bone mineral density (BMD) whether primary, e.g. osteogenesis imperfecta [2], or secondary e.g. acute lymphoblastic leukaemia, rheumatological conditions, Duchenne muscular dystrophy and glucocorticoid use [1, 3]. Moreover, children who have been identified with vertebral fractures, especially those with osteogenesis imperfecta and Duchenne muscular dystrophy are more likely to have multiple vertebral fractures [4, 5]. An early radiological diagnosis and accurate identification of patients with prevalent vertebral fracture is important for the effective targeting of therapy to prevent new fractures.

Currently, the gold standard for identifying vertebral fractures in children is the lateral spine radiograph. Recent studies, have shown that spine images acquired by dual energy X-ray absorptiometry (DXA) are comparable to radiographs [6-8], allowing reduced exposure to radiation. The diagnosis of vertebral fractures from DXA is termed vertebral fracture assessment (VFA).

There is no standardised technique for objective diagnosis of vertebral fractures in children, and clinical studies have shown that there is significant inter and intra-

observer variability in this population [3, 9-11]. Moreover, the limited studies carried out to assess morphometric analysis (MXA) using a 6-point semi-automated software programme in children have also shown poor observer reliability [8, 12].

The aim of this study, therefore, was to assess whether observer reliability and diagnostic accuracy of MXA for the identification of vertebral fracture in children would be improved by using a 33-point semi-automated programme compared to the 6-point programme.

4.5 Materials and methods

4.5.1 Study population

The study population included 100 (50 DXA and 50 radiographic (XR)) lateral spine images (performed on the same day on the same children) that were obtained as part of a larger prospective study involving 137 children; these children were recruited between November 2011 and February 2014 [6,12]. The sample selection was randomly made using Random Number Generator. All images belonged to patients recruited from a single centre. The majority of patients (80%) were those with suspected reduced BMD, e.g. osteogenesis imperfecta, inflammatory bowel disease, rheumatological conditions, and cystic fibrosis, attending the metabolic bone clinic for iDXA and lateral spine radiographs. Details of image acquisition have previously been reported [6]. The remaining 20% of patients were those attending spine clinics for suspected scoliosis.

4.5.2 Ethics statement

For the main study, approval of the Local Research Ethics Committee was sought and obtained, but was not separately required for this study. The study was registered with the local Research and Innovation Department prior to commencement.

4.5.3 Image analysis

XR and VFA images were independently evaluated for vertebral fracture by a research radiographer (R1) and an expert paediatric radiologist (R2), using two different semi-automated programmes (1) SpineAnalyzer™ (Optasia Medical, Cheadle, UK, version number: 4.0.2.19) and (2) AVERT™ (Optasia Medical, Cheadle, UK). SpineAnalyzer™ is Optasia's software based on an active appearance model. AVERT™ is partially derived from SpineAnalyzer™, but uses the latest appearance modelling technology (Random Forest regression voting constrained local models) from the University of Manchester software libraries (Table 4.1). Potentially, therefore, AVERT™ might be expected to provide more accurate fits [13].

Prior to commencing the study, R1 was trained to use the software programmes by a research associate in computing science and an expert radiologist (MSK research radiology fellow), learning from non-study spine images. In order to reduce observer bias, XR and VFA analyses were performed on different days, without accessing the subject's clinical information. Based on Buiang et al.'s study [14] which recommended that the minimum sample size required for conducting kappa agreement tests is between 10 and 30, repeat scoring was performed on 10 randomly selected patients blinded to previous reads.

Table 4.1 Comparison between AVERT™ and SpineAnalyzer programmes

	AVERT™	SpineAnalyzer™
Model	Based on Random Forest regression voting constrained local models (RFRV-CLM) [13]	Based on an active appearance model (AAM) [15]
Points	33-point semi-automated programme	6-point semi-automated programme
Modelling libraries	Optasia Medical libraries	The University of Manchester software libraries
Model modality	CT and DXA-VFA	X-ray or DXA-VFA
Posterior vertebral height calculation	Based on (posterior: posterior^{+2}, and posterior: posterior^{-2}) height ratios (+2 and -2 indicate the four neighbouring vertebrae the two immediately above [+2] and the two immediately below [-2] the vertebra of interest)	Based on (posterior: posterior^{+1}, and posterior: posterior^{-1}) height ratios (+1 and -1 indicate the vertebrae immediately above [+1] and below [-1] the vertebra of interest)
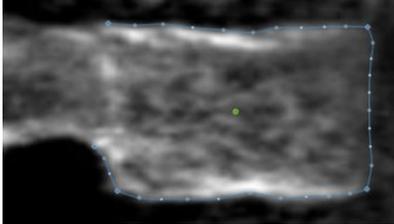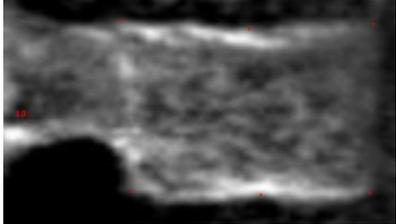
In line with the process associated with semi-automated analysis using SpineAnalyzer™, for each individual image (VFA or XR), the observer tracked T4 to L4 vertebral bodies by placing a single point at the centre of their centre (Figure 4.1a) and indicating to the software the highest identified vertebral body (for example, T4). Subsequently, the programme takes cognisance of all the identified vertebral bodies between T4 and L4 and automatically identifies 6-points that correspond to the midpoints of the superior and inferior endplates and the four corners of each vertebral body (Figure 4.1b); although these can be modified as necessary (Figure 4.1c). Importantly, the software does not recognise vertebral bodies above T4 or below L4, although unreadable vertebral bodies between these levels can be omitted from the readings. Once the 6 points have been placed, anterior, middle and posterior vertebral

heights are automatically determined by the software and, with the help of such measurements, the (anterior: posterior), (middle: posterior), (posterior: posterior[+1], and posterior: posterior[-1]) height ratios are calculated (+1 and -1 indicate the vertebrae immediately above [+1] and below [-1] the vertebra of interest). The vertebral bodies are then categorised according to the height loss ratio; height loss of 20%-25% (mild), height loss of 25%-40% (moderate) or height loss more than 40% (severe), based on the semi-quantitative scoring system developed by Genant [16].
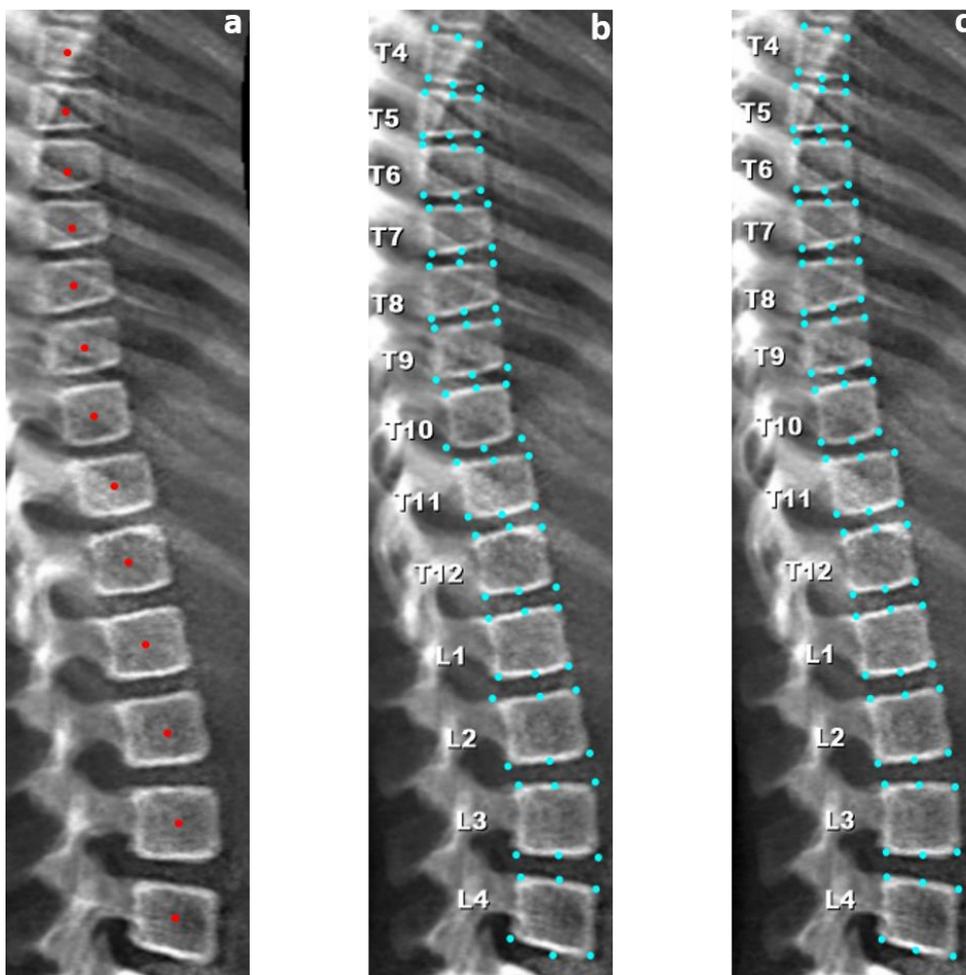
Figure 4.1 Analysing an iDXA lateral spine image using SpineAnalyzer™ (a) placement of a single point at the centre of each vertebral body (b) automatic 6-point annotation (c) manual correction of 6 points (e.g. anterior points of T10 and T12)

In the case of AVERT™, all lateral XR and VFA images (T4–L4) were analysed as follows: initial manual targeting of the centres of the vertebral bodies of interest (Figure 4.2a), then the software numbers the vertebral bodies accordingly. The software then automatically finds the positions of landmarks to enable a 33-point measurement (Figure 4.2b) for each vertebral body: 11 on the upper end-plate, 8 anteriorly, 11 on the lower end-plate, and 3 posteriorly. The software then allows these points to be moved by the observer, if deemed necessary, to correct any fitting failures (Figure 4.2c).
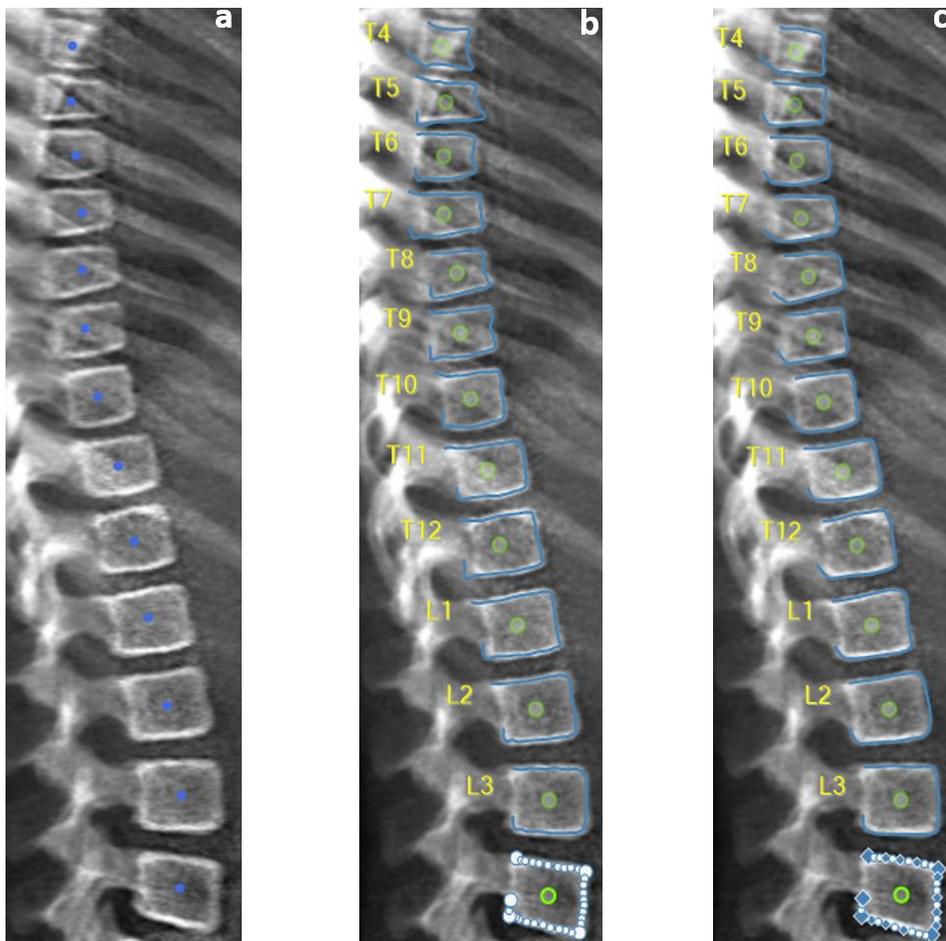
Figure 4.2 Analysing an iDXA lateral spine image using AVERT™ (a) placement of a single point at the centre of each vertebral body (b) automatic 33-point annotation (c) manual correction of 33 points at L4

Subsequently, the confirmed points are used by the software to calculate the anterior, middle and posterior vertebral heights, which are used for the determination of the shape of any deformity. From these measurements, the (anterior: posterior), (middle: posterior), (posterior: posterior^{+2}, and posterior: posterior^{-2}) height ratios are calculated (+2 and -2 indicate the four neighbouring vertebrae the two immediately above [+2] and the two immediately below [-2] the vertebra of interest). Thereafter, the vertebral bodies are classified as per their height ratios, on the basis of Genant's scoring system [16].

For this study, in terms of identifying vertebral levels, the first vertebral body that was not associated with a pair of ribs was marked as L1, with the lowermost vertebral body associated with ribs then marked as T12.

For both programmes, the operator is able to move the points for improved fit to vertebral shape. The time to conduct MXA for both programmes was measured for R1 and R2 on 20 randomly selected images.

4.5.4 Statistical analysis

SPSS statistics software version 24 (IBM, Armonk, NY, USA) and Microsoft® Excel 2016 were employed for data analysis. The reference standard for diagnostic accuracy (sensitivity, specificity, false positive (FP) and false negative (FN) rates) calculations was taken from a previous consensus reached by three paediatric radiologists using a simplified ABQ (sABQ) scoring system [11]. For these calculations of diagnostic accuracy, all sABQ, SpineAnalyzer™ and AVERT™ scores of 0 or 1 were interpreted as, "no clinically significant fracture". Inter and intraobserver agreement were calculated using Cohen's kappa with 95% confidence interval [CI].

4.6 Results

The mean age of the 50 subjects at the time of image acquisition was 9.6 years (range 5 to 15) and 21 (42%) were male.

According to the reference standard, 34 (68%) had at least one fracture. Amongst these 34 patients, there was a total of 175 vertebral fractures, 132 (75%) were mild, 41 (23%) were moderate and 2 (1%) were severe. Only 2 of the 34 patients (4%) had severe fractures.

A total of 2600 individual vertebral bodies (T4–L4) collated from both radiographs and VFA were assessed by each observer using SpineAnalyzer™ and AVERT™.

All vertebral fracture locations were distributed throughout the thoracic and lumbar spine. The total number and severity of vertebral fractures identified through each technique is shown in Table 4.2. In general, the number and severity of vertebral fractures at both subject and vertebral levels varied between the gold standard and the four investigated methods, however, the severity of vertebral fractures was similar for XR and VFA when using AVERT™. Both methods identified slightly more mild fractures compared to moderate or severe fractures for both observers irrespective of image modality.

Table 4.2 Prevalence (%) of vertebral fractures in study cohort (n = 50, 650 vertebrae) at vertebral and subject levels

	DXA AVERT™				XR AVERT™				DXA SpineAnalyzer™				XR SpineAnalyzer™				Reference Standard	
	Per vertebra		Per subject		Per vertebra		Per subject		Per vertebra		Per subject		Per vertebra		Per subject		Per vertebra	Per subject
	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2		
No fracture	554 (85%)	502 (77%)	18 (36%)	10 (20%)	561 (86%)	557 (85%)	18 (64%)	19 (38%)	558 (86%)	611 (94%)	17 (34%)	30 (60%)	597 (92%)	616 (95%)	30 (60%)	37 (74%)	475 (73%)	16 (32%)
At least one mild fracture (≤ 24% loss of height)	59 (9%)	85 (13%)	28 (56%)	37 (74%)	48 (7%)	47 (7%)	29 (58%)	27 (54%)	56 (9%)	26 (4%)	31 (62%)	16 (32%)	23 (4%)	17 (3%)	16 (32%)	11 (22%)	132 (20%)	31 (62%)
At least one moderate fracture (25% to 40% loss of height)	22 (3%)	51 (8%)	15 (30%)	26 (52%)	27 (4%)	37 (6%)	17 (34%)	14 (28%)	24 (4%)	12 (2%)	13 (26%)	7 (14%)	16 (2%)	9 (1%)	11 (22%)	6 (12%)	41 (6%)	8 (16%)
At least one severe fracture (≥ 41% loss of height)	15 (2%)	12 (2%)	4 (8%)	5 (10%)	14 (2%)	9 (1%)	4 (8%)	4 (8%)	12 (2%)	1 (<1%)	4 (8%)	1 (2%)	14 (2%)	8 (1%)	3 (6%)	3 (6%)	2 (0.3%)	2 (4%)

Sensitivity and specificity of AVERT™ and SpineAnalyzer™ per vertebral level for both modalities (DXA and XR) for all vertebrae from T4 to L4 are shown in Figures 4.3 and 4.4 respectively.
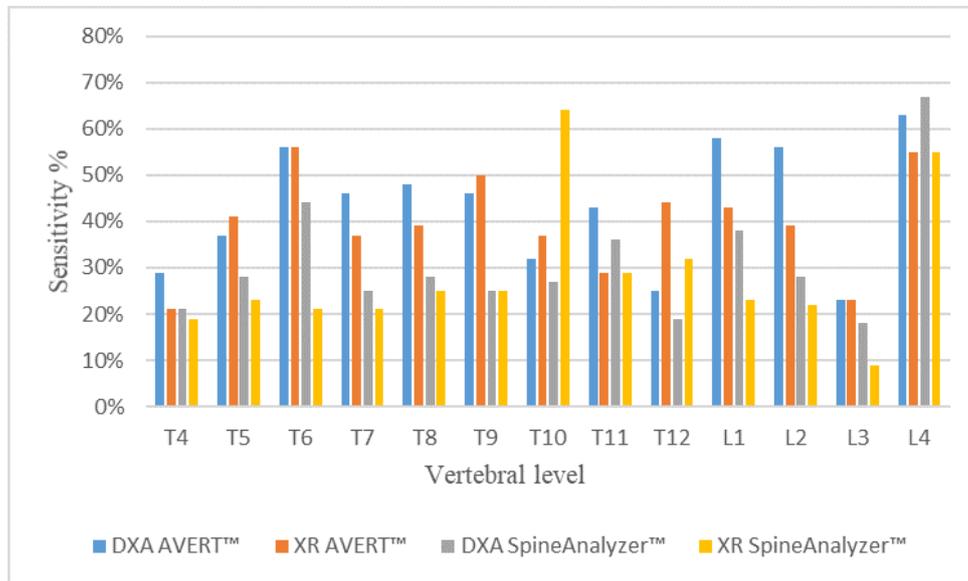
Figure 4.3 Sensitivity identified for all techniques per vertebral level against the 'gold standard' (consensus read by three experienced paediatric radiologists using spine radiographs)
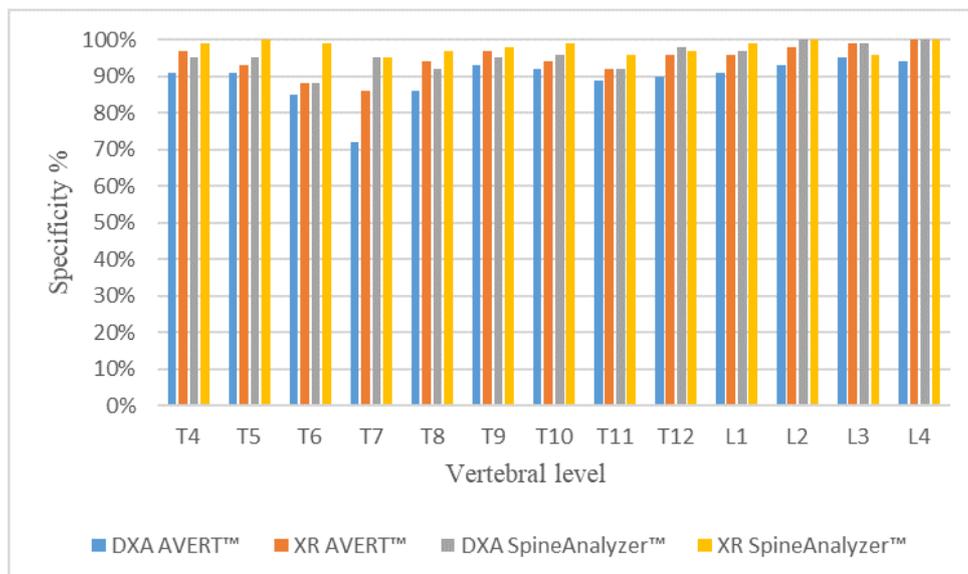
Figure 4.4 Specificity identified for all techniques per vertebral level against the 'gold standard' (consensus read by three experienced paediatric radiologists using spine radiographs)

Sensitivity, specificity, reliability (kappa- 95 % CI) and FN and FP rates of

SpineAnalyzer™ and AVERT™ for both modalities are summarised in Table 4.3.

Table 4.3 Diagnostic accuracy of AVERT™ and SpineAnalyzer™ for vertebral fracture diagnosis in children

	DXA AVERT™	XR AVERT™	DXA SpineAnalyzer™	XR SpineAnalyzer™
Sensitivity (%)	41	36	31	26
Specificity (%)	91	95	96	98
False negative (%)	59	64	69	75
False positive (%)	9	5	4	2
Agreement (%)	79	80	79	78
Kappa (95 % CI)	0.34 (0.26, 0.40)	0.37 (0.27, 0.46)	0.31 (0.21, 0.44)	0.26 (0.16, 0.35)

Figure 4.5 shows the agreement between the two programmes for a selected DXA image. Overall, there was fair agreement (assessed by kappa statistics) between the four techniques and the consensus evaluation in terms of identifying vertebral fracture: the average kappa score ranged from 0.26 to 0.37 (95% CI: 0.16, 0.46), with XR SpineAnalyzer™ having the lowest score 0.26 (95% CI: 0.16, 0.35) and XR AVERT™ having the highest score 0.37 (95% CI: 0.27, 0.46). However, no statistically significant differences were noticed between all the techniques assessed.
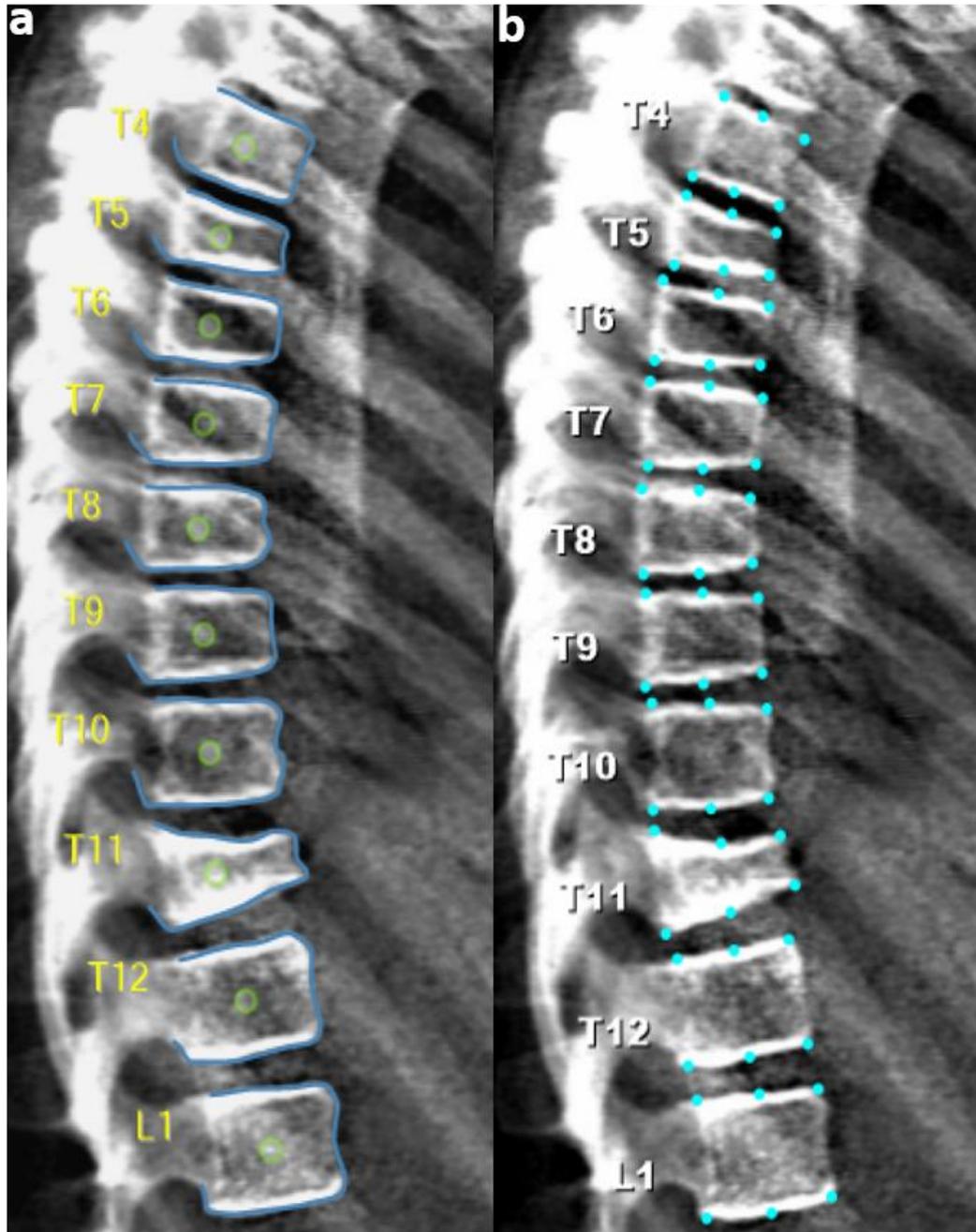
Figure 4.5 10-year-old child with osteogenesis imperfecta. Lateral spine DXA image analysed by R1 using AVERT™ (a) and SpineAnalyzer™ (b) which illustrates:

- Agreement: Both programmes identified a severe fracture at T11; moderate fractures at T5 and T6; mild fractures at T7 and T8

- Disagreement: T9 identified as mild fracture by AVERT™ but normal by SpineAnalyzer™

- Gold standard values: T5, T7, T8 and T9 classified as mild fractures; T6 as normal and T11 as a moderate fracture

Table 4.4 summarises inter and intraobserver agreement of all four methods for the two observers. There was moderate interobserver agreement for all methods, with kappa ranging from 0·41 to 0·47 (95% CI: 0·25, 0·66). In contrast, intraobserver agreement ranged from moderate to good, with mean kappa values for R1 and R2 ranging from 0·50 to 0.79 and 0.59 to 0.78 respectively; SpineAnalyzer™ XR had the lowest score for both observers. For AVERT™, kappa scores for R1 and R2 on DXA scans were 0.79 (95% CI: 0.69, 0.90) and 0.73 (95% CI: 0.66, 0.82) respectively.

Table 4.4 Summary of inter and intraobserver agreement for all methods

	Method		Observer	Kappa			Agreement (%)
	Software	*Modality*		**Mean**	**Min**	**Max**	
Interobserver Agreement	AVERT™	DXA	R1 vs R2	0·47	0·27	0·66	95
	AVERT™	Radiographs	R1 vs R2	0·46	0·21	0·77	94
	SpineAnalyzer™	DXA	R1 vs R2	0·41	0·25	0·65	96
	SpineAnalyzer™	Radiographs	R1 vs R2	0·42	0·14	0·73	95
Intraobserver Agreement	AVERT™	DXA	R1	0.79	0.57	1.00	98
			R2	0.73	0.41	1.00	97
		Radiographs	R1	0.78	0.57	1.00	98
			R2	0.77	0.34	1.00	98
	SpineAnalyzer™	DXA	R1	0.66	0.34	1.00	94
			R2	0.78	0.54	1.00	97
		Radiographs	R1	0.50	0.30	0.69	91
			R2	0.59	0.41	1.00	92

Table 4.5 summarises the overall results of this current study and compares with those of all previous studies that have evaluated semi-automated software techniques in children [7, 8, 12, 17].

Table 4.5 Summary of diagnostic accuracy and observer agreement of semi-automated software techniques in children

Study	Gold Standard (Radiographs)	Method		Sensitivity (%)	Specificity (%)	False negative rate (%)	False positive rate (%)	Interobserver agreement (kappa)	Intraobserver agreement (kappa)
		Software	*Modality*						
Kyriakou et al [7]	Non radiologist reader	Six-point analysis	DXA	75	98	-	-	0.79	-
Crabtree et al [8]	An expert paediatric radiologist	Six-point analysis	DXA	79	71	3	25	0.32	-
Alqahtani et al [12]	Consensus arrived by three paediatric radiologists	Six-point analysis (SpineAnalyzer™)	Radiographs	18	97	-	-	0.05-0.47	0.25-0.61
Diacinti et al [17]	Consensus of two skeletal radiologists	Six-point analysis (the Hologic QDR Physician's Viewer software (version 7.02)	DXA	66	95	59	10	0.71	-
Current study	Consensus arrived by three paediatric radiologists	Six-point analysis (SpineAnalyzer™)	DXA	31	96	69	4	0.41	0.73-0.79
			Radiographs	26	98	75	2	0.42	0.50-0.59
Current study	Consensus arrived by three paediatric radiologists	33-point analysis (AVERT™)	DXA	41	91	59	9	0.47	0.73-0.79
			Radiographs	36	95	64	5	0.46	0.77-0.78

The time taken by R1 and R2 per image/patient averaged 8±4 minutes (range, 6–14) and 6 ±2.0 minutes (range, 4–9 minutes) respectively for AVERT™ and 6±2 minutes (range, 3–10) and 3 ±1 minute (range 2–7 min) respectively for SpineAnalyzer™.

4.7 Discussion

According to the ISCD criteria, the definition of osteoporosis in children is dependent on the identification of one or more vertebral fractures. In the absence of vertebral fractures the diagnosis may be made depending on the presence of a bone mineral density Z-score of ≤ -2.0, as well as the number of long bone fractures sustained by the ages of 10 (≥ 2) and 19 (≥ 3) years (1). It is therefore important to diagnose vertebral fractures in children at an early stage to allow appropriate treatment plans to be established, such as bisphosphonates, which treat existing fractures as well as reduce the risk of future fractures [18].

Although there are several commercially available programmes for quantitative vertebral morphometry assessment in adults, there is as yet no specific semi-automated software for children. In adult subjects, the agreement between observers using 6-point technique programmes e.g. SpineAnalyzer™ (Optasia Medical, Cheadle, UK) and MorphoXpress (MorphoXpress, P&G Pharmaceuticals, Rusham Park, Egham, UK) has been reported to be higher than that in this study [19-23]. These previous studies show that the 6-point technique programmes have very high sensitivity and specificity, reaching 98% and 99% respectively, and excellent interobserver agreement of 99%, with kappa ranging from 0.86 to 0.97. In fact, these adult studies show significantly higher diagnostic accuracy than those of all previous studies evaluating 6-point semi-automated programmes in children [7, 8, 12, and 17].

The purpose of this current study therefore was to ascertain whether observer reliability and diagnostic accuracy of MXA for the identification of vertebral fractures in children would be improved by using a 33-point semi-automated programme compared to the 6-point programme for either VFA or radiographs. We analysed images from 50 subjects used for a previous study [12]. To our knowledge, this is the first report to assess two programmes on two different modalities (VFA and radiographs) for the identification of vertebral fractures in children.

Compared to the consensus reached by the three radiology experts, the overall sensitivity of the 6- and 33- point semi-automated techniques ranged from 26% to 31% and 36% to 41% respectively. These results are slightly higher than the results from a previous study, in which five readers with different levels of experience assessed the same version of the SpineAnalyzer™ software on 137 radiographs and showed overall sensitivity of only 18% (95% CI: 14-2), while overall specificity was 97% (95% CI: 97-98) [12]. The 50 images used in the current study were randomly selected from the 137 used in [12], and showed improved overall sensitivity and specificity for SpineAnalyzer™ of 26% to 31% and 96% to 98% respectively; and 36% to 41% and 91% to 95% respectively for AVERT™.

In the current study, diagnostic accuracy parameters for both software programmes was somewhat comparable to previous studies [7, 8, 17] (Table 4.5). The current study has the strength of using a consensus read by three paediatric radiologists, each with minimum 13 years' experience, as the reference standard.

We have demonstrated that MXA on DXA images is comparable to MXA on radiographs for identifying clinically significant osteoporotic fractures irrespective of the software programme. However, MXA has low diagnostic accuracy and poor

observer reliability, with high FN and FP rates. Both programmes underdiagnosed the prevalence of mild fractures; of the 132 reference standard mild vertebral fractures, only 59, 48, 56 and 23 were identified by DXA AVERT™, XR AVERT™, DXA SpineAnalyzer™ and XR SpineAnalyzer™ by R1 respectively and 85, 47, 26 and 17 by R2 respectively. Moderate and severe vertebral fractures ($\geq$25% loss of height in the vertebral body) are readily identified by the naked eye, it is the detection of mild fractures that is clinically problematic [8]. Far from improving the detection of mild fractures, it would seem that MXA underdiagnoses them.

Despite the limitation of the increased reading time associated with AVERT™, it showed slightly higher accuracy for diagnosis of vertebral fractures in children compared to SpineAnalyzer™. However, for both programmes the time was longer in subjects with moderate and/or severe vertebral fractures compared to those with no fracture. As the FP and FN rates were high in this study, it can be said that the target of FP and FN rate that would make a test suitable for clinical use is a clinical decision, but 5% might be acceptable. Moreover, it is easy enough to compare two percentages using standard methods.

The poor observer reliability for both programmes may have some explanations. First, there is an inherent subjectivity related to the semi-automated placement of points. Since the placement of these points still relies heavily on the experience of the observer, the correct location of the points may be problematic. Secondly, both programmes use the Genant system as their reference, which bases the assessment only on the loss of height of vertebral bodies, while the gold standard uses the sABQ method, which is a visual method that takes account of alterations in the vertebral endplates which may be non-fracture related. Currently, the authors believe that visual methods such as the sABQ approach are more accurate methods of assessing vertebral fractures in children.

4.8 Conclusion

Our results show that AVERT™ has a slightly higher accuracy for diagnosis of vertebral fractures in children compared to SpineAnalyzer™; but both methods have low diagnostic accuracy and observer reliability and we conclude that until the software programmes have been specifically improved, MXA cannot be used as a diagnostic tool for vertebral fracture diagnosis in children.

4.9 Acknowledgments

4.10 References

1. Bishop N, Arundel P, Clark E, et al (2014) Fracture Prediction and the Definition of Osteoporosis in Children and Adolescents: The ISCD 2013 Pediatric Official Positions J Clin Densitom 17:275-280

2. Bishop N (2010) Characterising and treating osteogenesis imperfecta. Early Hum Dev 86:743-746

3. Halton J, Gaboury I, Grant R, et al (2009) Advanced Vertebral Fracture Among Newly Diagnosed Children With Acute Lymphoblastic Leukemia: Results of the Canadian Steroid-Associated Osteoporosis in the Pediatric Population (STOPP) Research Program. J Bone Miner Res 24:1326-1334

4. Alos N, Grant R, Ramsay T, et al (2012) High incidence of vertebral fractures in children with acute lymphoblastic leukemia 12 months after the initiation of therapy. J Clin Oncol 30:2760

5. Bothwell JE, Gordon KE, Dooley JM, Mac Sween J, Cummings EA, Salisbury S (2003) Vertebral Fractures in Boys with Duchenne Muscular Dystrophy. Clin Pediatr (Phila) 42:353-356

6. Adiotomre E, Summers L, Allison A, et al (2017) Diagnostic accuracy of DXA compared to conventional spine radiographs for the detection of vertebral fractures in children. Eur Radiol 27:2188-2199.

7. Kyriakou A, Shepherd S, Mason A, Ahmed SF (2015) A critical appraisal of vertebral fracture assessment in paediatrics Bone 81:255-259

8. Crabtree N, Chapman S, Högler W, et al (2017) Vertebral fractures assessment in children: Evaluation of DXA imaging versus conventional spine radiography Bone 97: 168-174

9. Mayranpaa MK, Helenius I, Valta H, Mayranpaa MI, Toiviainen-Salo S, Makitie O (2007) Bone densitometry in the diagnosis of vertebral fractures in children: Accuracy of vertebral fracture assessment. Bone 41:353-359

10. Siminoski K, Lentle B, Matzinger MA, Shenouda N, Ward LM, Canadian SC (2014) Observer agreement in pediatric semiquantitative vertebral fracture diagnosis Pediatr Radiol 44:457-466

11. Adiotomre E, Summers L, Allison A, et al (2016) Diagnosis of vertebral fractures in children: is a simplified algorithm-based qualitative technique reliable? Pediatr Radiol 46:680-688

12. Alqahtani FF, Messina F, Kruger E, et al (2017) Evaluation of a semi-automated software program for the identification of vertebral fractures in children. Clin Radiol 72:904.e11-904.e20

13. Lindner C, Bromiley PA, Ionita MC, Cootes TF (2015) Robust and accurate shape model matching using random forest regression-voting. IEEE Trans Pattern Anal Mach Intell 37:1862-1874

14. Bujang, M.A. and Baharum, N. (2017) Guidelines of the minimum sample size requirements for Cohen's Kappa. Epidemiol Biostat Public Health 14:e12267-12261

15. Cootes, T.F., Edwards, G.J. and Taylor, C.J. (2001) Active appearance models. IEEE Trans Pattern Anal Mach Intell 6:681-685

16. Genant HK, Wu CY, Vankuijk C, Nevitt MC (1993) Vertebral fracture assessment using a semiquantitative technique. J Bone Miner Re*s* 8:1137-1148

17. Diacinti D, Pisani D, D'Avanzo M, et al (2015) Reliability of Vertebral Fractures Assessment (VFA) in Children with Osteogenesis Imperfecta. Calcif Tissue Int 96:307-312

18. Shaw NJ (2008) Management of osteoporosis in children. *Eur J Endocrinol* 59:S33-S39

19. Guglielmi G, Palmieri F, Placentino MG, D'Errico F, Stoppino LP (2009) Assessment of osteoporotic vertebral fractures using specialized workflow software for 6-point morphometry. Eur J Radiol 70:142-148

20. Guglielmi G, Stoppino LP, Placentino MG, D'Errico F, Palmieri F (2009) Reproducibility of a semi-automatic method for 6-point vertebral morphometry in a multi-centre trial. Eur J Radiol 69:173-178

21. Sanfélix-Genovés J, Arana E, Sanfélix-Gimeno G, Peiró S, Graells-Ferrer M, Vega-Martínez M (2012) Agreement between semi-automatic radiographic morphometry and Genant semi-quantitative method in the assessment of vertebral fractures. Osteoporos Int 23:2129-2134

22. Birch C, Knapp K, Hopkins S, Gallimore S, Rock B (2015) SpineAnalyzer (TM) is an accurate and precise method of vertebral fracture detection and classification on dual-energy lateral vertebral assessment scans. Radiography 21:278-281

23. Diacinti D, Del Fiacco R, Pisani D, et al (2012) Diagnostic Performance of Vertebral Fracture Assessment by the Lunar iDXA Scanner Compared to Conventional Radiography. Calcif Tissue Int 91:335-342

Chapter Five

5. Diagnostic performance of morphometric vertebral fracture analysis (MXA) in children using a 33-point software programme

Fawaz F Alqahtani[1,2], Nicola J Crabtree[3], Paul A. Bromiley[4], Timothy F. Cootes[4], Penny Broadley[5], Isla Lang[5], Amaka C Offiah[1,5]

Correspondence: Dr. Amaka C Offiah

University of Sheffield

Academic Unit of Child Health

Damer Street Building

Sheffield Children's NHS Foundation Trust

Western Bank

Sheffield

S10 2TH

Phone: 0114 271 7557

E–mail: a.offiah@sheffield.ac.uk

1 Academic Unit of Child Health, University of Sheffield, Sheffield, UK

2 Department of Radiological Sciences, College of Applied Medical Sciences, Najran University, Najran, Kingdom of Saudi Arabia

3 Department of Endocrinology and Diabetes, Birmingham Children's Hospital, Birmingham, UK

4 Division of Informatics, Imaging and Data Sciences, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

5 Radiology Department, Sheffield Children's NHS Foundation Trust, Sheffield, UK

5.1 Abstract

Background

There is significant inter and intraobserver variability in diagnosing vertebral fractures in children. We aimed to evaluate the diagnostic accuracy of morphometric vertebral fracture analysis (MXA) using AVERT™ (Optasia Medical, Cheadle, UK), a software programme designed for adults, on dual-energy X-ray absorptiometry (DXA) images of children.

Methods

Lateral spine DXA images of 420 children aged between 5 and 18 years were retrospectively reviewed. Vertebral fracture assessment (VFA) by an expert paediatric radiologist using Genant's semiquantitative scoring system served as the gold standard. All 420 DXA scans were analysed by a trained radiographer, using semi-automated software (33-point morphometry). VFA of a random sample of 100 images was performed by an experienced paediatric clinical scientist. MXA of a random sample of 30 images were analysed by three paediatric radiologists and the paediatric clinical scientist. Diagnostic accuracy and inter and intraobserver agreement (kappa statistics) were calculated.

Findings

Overall sensitivity, specificity, false positive (FP) and false negative (FN) rates for the radiographer using the MXA software were 80%, 90%, 10%, and 20% respectively and for mild fractures alone were 46%, 92%, 8%, and 54% respectively. Overall sensitivity, specificity, FP, and FN rates for the four additional observers using MXA were 89%, 79%, 21%, and 11% respectively and for mild fractures alone were 36%, 86%, 14%, and 64% respectively. Agreement between two expert observers was fair to good for

VFA and MXA [kappa = 0·29 to 0·76 (95% CI: 0·17 – 0·88) and 0·29 to 0·69 (95% CI: 0·17 – 0·83)] respectively.

Interpretation

MXA using a 33-point technique developed for adults is not a reliable method for the identification of mild vertebral fractures in children. A paediatric standard is required which not only incorporates specific vertebral body height ratios but also the age-related physiological changes in vertebral shape that occur throughout childhood.

Funding

5.2 Introduction

Osteoporotic fractures may occur in children and adolescents with low bone mineral density (BMD) either as a primary disorder (e.g., osteogenesis imperfecta),[1] or secondary to various disorders and medications including acute lymphoblastic leukaemia, rheumatic disorders, inflammatory bowel disease, Duchenne muscular dystrophy, and glucocorticoid therapy [2, 3].

Vertebral fractures represent a significant proportion of all osteoporotic fractures and thus, given a lack of major trauma or local disease, presentation with one or multiple vertebral fractures is a strong indicator of low bone mass in children [4]. Most vertebral fractures are not identified clinically, which may be problematic, given the high levels of morbidity they may be associated with. Children with chronic low BMD are screened at least annually to assess prevalent and incident vertebral fractures, thus a method of accurate detection of these fractures must be devised to allow prompt therapeutic interventions.

Until very recently, lateral spine radiographs were the main method for identifying vertebral fractures. However, the latest bone densitometers have made it possible to conduct vertebral fracture assessment (VFA) from dual-energy X-ray absorptiometry (DXA) scans. This technique is considered preferable due to similar sensitivity and specificity when compared to radiographs, as well as the advantage of reduced radiation dose [5–8]. The available scoring systems for VFA in adults have also been evaluated for utilisation in children: these systems include Genant's semiquantitative technique (SQ);[2, 7, 9] the algorithm–based qualitative (ABQ) technique,[10] and software programmes that allow morphometric analysis (MXA) [6–8, 11]. Results have been variable, with the largest studies showing low diagnostic accuracy of VFA and MXA,

135

particularly for mild fractures, which are the most important to detect in order to prevent progression [5, 6, 11].

The newest generation of bone densitometers are capable of enhancing the diagnostic utility of DXA through integration with semi–automated software that helps to diagnose vertebral fractures. In terms of recent refinements to MXA, the shape-based statistical modelling technique for semi-automated quantitative morphometry has been devised for detection of fractures in adults,[12] and this technical development may also improve analysis in children in terms of efficiency and accuracy.

The aim of this study was to evaluate the diagnostic accuracy of MXA through the use of a novel semi-automated 33-point morphometric software tool, AVERT™ (Optasia Medical, Cheadle, UK), in a cohort of children with chronic disease, using the latest iDXA imaging technology in the hands of various observers compared to the reference standard of a visual SQ method applied by an experienced paediatric radiologist.

5.3 Materials and methods

5.3.1 Study population

The Picture Archiving and Communication System of Sheffield Children's Hospital (SCH) was searched for all lateral spine iDXA images performed between November 2011 and November 2016 in children aged between five and 18 years old. The total of 2800 images was divided into yearly cohorts based on age and 15 lateral spine iDXA images were randomly selected for each year of age and both sexes, giving a total of 420 iDXA lateral spine images which were anonymised and included in the study. Bone mineral density (BMD) for both lumbar spine (L2–L4) and total body less head (TBLD) were performed as part of the same investigation.

5.3.2 Ethics statement

The study protocol was approved by the Local Health Research Authority (HRA reference number: 210524). The study was also registered with the local Research and Innovation Department and conducted in accordance with the Declaration of Helsinki and the NHS Research Governance Framework.

5.3.3 Lateral spine imaging

Lateral spine DXA scans were acquired using a Lunar GE iDXA machine (GE Healthcare Lunar iDXA, Buckinghamshire, UK), following the manufacturer's recommendations. Briefly, the child was positioned in the decubitus position on the scanning table, with their knees flexed upwards towards the chest, so that the spine was parallel to the table with their arms above their head and away from the area to be scanned. Foam padding was used to obtain and maintain the required position.

5.3.4 Image analysis

All images were analysed using AVERT™ (Optasia Medical, Cheadle, UK). AVERT™ is a software programme based on a 33-point morphometric technique and uses the latest appearance modelling technology (Random Forest regression voting constrained local models) developed by the University of Manchester [13]. Figure 5.1 is a flow chart of the reporting pathway described below.

5.3.4.1 Reference standard (420 VFA, R1)

For the reference standard, identification of vertebral fractures was performed on the 420 VFA by visually assessing the T4 to L4 vertebrae, relying on an experienced paediatric radiologist (R1) i.e., with no software involved, as is the current clinical standard. The previous two studies used the ABQ method, while this study used the Genant method as requested by the computing department at the University of

Manchester (UoM) to be able to use the same categorisation of AVERT™ programme. In other words, as the software programme is not able to identify fractures when height loss is below 20%, using the Genant method may help to reduce false positive and negative ratios. Quantitative measurements only took place at the reader's discretion. Vertebrae were categorised by this visual semi–quantitative method as 0 "non-fractured", 1 "mild fracture", 2 "moderate fracture", and 3 "severe fracture" based on Genant et al. classification [14]. Grades 0, 1, 2, and 3 entail loss of height of ≤20%, 21% to 25%, 26% to 40%, and ≥ 41% respectively. Vertebral fractures are manifested by a variety of alterations in shape, including "wedge", "biconcavity", or "crush", depending on the site of maximum reduction in vertebral height (anterior, middle, or generalised respectively). Additionally, vertebrae felt to be fractured but with ≤ 20% reduction in height and vertebrae with loss of height felt to represent physiological wedging were reported by the radiologist.

For consistency of vertebral level detection between observers, prior to study commencement at the stage of image anonymisation, R2 placed a marker at T12 for all images, confirmed by R1. The lowermost vertebral body associated with a pair of ribs was always designated as T12.

5.3.4.2 Diagnostic accuracy of MXA (420 iDXA, R2)

A radiographer (R2) used AVERT™ to perform MXA on the 420 selected DXA images. Prior to commencing the study, the radiographer was trained to use the software programme by experts from the University of Manchester, who participated in developing the software (the training was provided by a research associate in imaging sciences and an expert radiologist, using 72 non-study spine images).

5.3.4.2 Intraobserver agreement of MXA (100 iDXA, R2)

To evaluate intraobserver agreement of MXA for R2, DXA images of 100 subjects were randomly selected from the study population for a second read. In order to reduce recall bias, the repeat scoring was performed after an interval of approximately 30 days.

5.3.4.3 Observer agreement of MXA (30 iDXA, R1, R3, R4, R5)

To ascertain observer agreement of MXA more widely, three consultant paediatric musculoskeletal radiologists (R1, R4, R5), each with a minimum of 13 years' experience, and an experienced clinical scientist (R3), independently performed MXA on 30 iDXA images randomly selected from the 100 interpreted by R2. Following an interval of at least 2 weeks, 10 of the 30 iDXA images were randomly selected for a second read by the same four observers to allow calculation of intraobserver agreement of MXA.

5.3.4.4 Interobserver agreement of VFA (100 iDXA, R1, R3)

To evaluate interobserver agreement of VFA, an experienced paediatric clinical scientist (R3) independently used the SQ grading scale for visual assessment (VFA) of the same 100 iDXA used for R2's second read. The results were compared to the reference standard to assess interobserver agreement of VFA.

Sensitivity, specificity, false positive (FP), and false negative (FN) rates were calculated for all grades of fracture and for mild fractures alone.
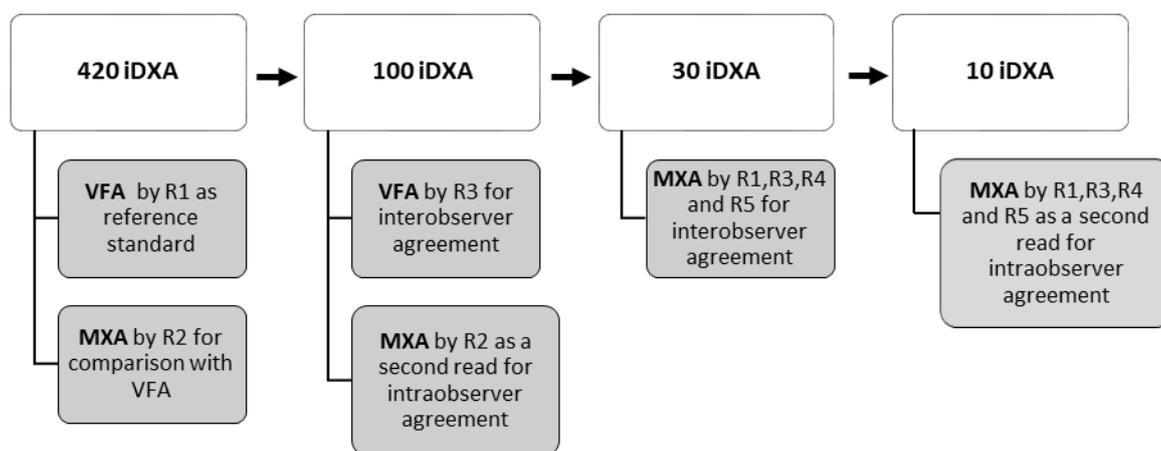
Figure 5.1 Flow chart summarising the reporting pathway

5.3.5 Morphometric analysis

The first step in MXA required the observers to identify all vertebrae from T4 to L4 by manually placing a single point at the centre of each vertebral body, then the software identified the vertebral bodies accordingly (i.e. T4 as the highest and L4 as the lowest vertebra) (Figure 5.2). Subsequently, the programme automatically outlined each labelled vertebra with 33 measurement points: eleven on the upper end-plate, eight on the anterior margin, eleven on the lower end-plate, and three on the posterior margin (leading to 33 points for each vertebral body). The observers reviewed the images and, if necessary, modified these points. From these confirmed points, the software then computed the anterior, middle and posterior (h^a, h^m and h^p) heights and calculated the wedge ratio (h^a/h^p), biconcave ratio (h^m/h^p) and crush ratio (h^p/h^{p+2} or h^p/h^{p-2}), where +2 and -2 indicate the four neighbouring vertebrae, i.e. the two immediately above [+2]

and the two immediately below [-2] the vertebra under examination. Based on the semi–quantitative scoring system developed by Genant et al., vertebrae were classified according to their height loss ratios as normal or mild, moderate, or severe fracture for height loss of $< 21\%$, 21%–25%, 26%–40% and $\geq 41\%$ respectively.
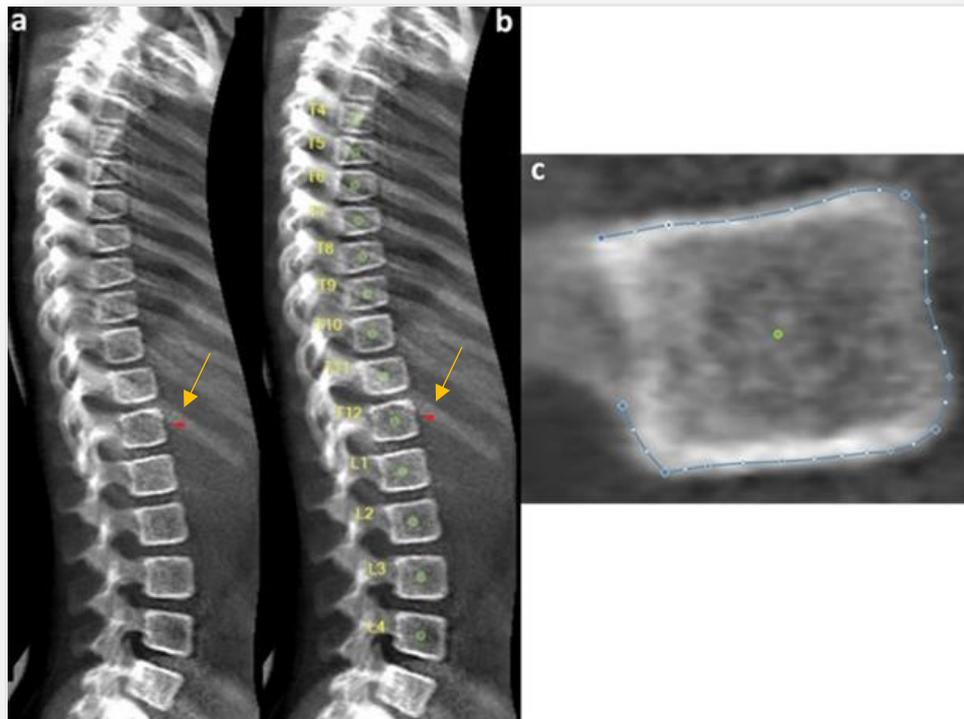
Figure 5.2 Technique used to perform semi-automated quantitative morphometric measurements (AVERT™). a) Lateral iDXA scan of the entire spine of a 9 -year-old female with osteogenesis imperfecta; b) identified vertebral bodies from T4 to L4; c) 33 points placed to outline T12. The arrow points to the T12 marker that ensured consistency between readers for vertebral level identification (lowest vertebral body associated with a rib).

5.3.6 Statistical analysis

We report demographic and bone densitometry data (bone mineral density (BMD, g/cm3) and z-score for both L2–L4 and TBLD). The frequency of vertebral fracture severity for each observer and for all vertebrae from T4 to L4 was calculated. Inter and intraobserver agreement and associated 95% confidence intervals (CI) were calculated using the kappa statistic. Diagnostic accuracy of observers (sensitivity, specificity, FP

and FN rates) was calculated. Analyses were performed both at the subject and at the individual vertebral level. We analysed prevalent vertebral fractures in three groupings: (1) any fracture (mild, moderate and severe); (2) moderate and severe, and (3) mild fracture.

Statistical analyses were conducted using SPSS statistics software version 24 (IBM, Armonk, NY, USA) and Microsoft® Excel 2016.

5.4 Results

We included 420 lateral iDXA scans in children aged between 5 and 18 years (30 per year of age being the typical number used to train software); 210 (50 %) were male; 380 (90%) had osteogenesis imperfecta, 12 (3%) Duchenne muscular dystrophy, 8 (2%) polyostotic fibrous dysplasia, and 20 (5%) other conditions including anorexia nervosa, diabetes mellitus, juvenile dermatomyositis and coeliac disease. Descriptive and clinical data are presented in Table 5.1.

Table 5.1 Summary of demographic and bone densitometry data of study subjects (mean and SD), n = 420*

Age (Year)	Height (cm)	SD	Weight (kg)	SD	L2–L4 BMD (g/cm^3)	SD	L2–L4 BMD (z–score)	SD	TBLH BMD (g/cm^3)	SD	TBLH BMD (z–score)	SD
5	105·5	10·85	17·42	3·30	0·559	0·13	-0·857	1·79	0·489	0·16	-0·687	0·72
6	111·85	8·90	20·07	3·76	0·616	0·10	-0·739	1·35	0·601	0·08	-0·478	0·90
7	116·77	13·57	23·91	6·28	0·669	0·11	-0·306	1·37	0·661	0·07	-0·106	1·19
8	126·37	8·53	30·44	11·39	0·735	0·24	-0·120	1·75	0·712	0·13	0·093	1·48
9	133·31	12·68	32·95	13·38	0·668	0·07	-0·736	1·05	0·659	0·11	-0·927	1·17
10	138·04	8·23	35·12	8·62	0·737	0·13	-0·507	1·35	0·763	0·10	-0·433	1·11
11	142·23	10·99	38·75	16·78	0·867	0·33	-0·538	1·40	0·802	0·12	-0·377	1·13
12	143·72	11·73	40·30	10·75	0·824	0·12	-0·908	0·80	0·805	0·13	-0·758	0·97
13	156·17	9·46	47·26	10·95	0·863	0·35	-0·127	1·79	0·902	0·16	-0·227	1·53
14	158·34	11·09	52·34	20·75	0·948	0·23	-0·360	1·51	0·966	0·12	-0·260	1·21
15	160·21	9·90	49·74	5·14	1·080	0·20	-0·442	1·39	0·970	0·11	-0·567	0·92
16	161·49	6·08	60·91	13·31	1·147	0·23	-0·287	1·77	0·969	0·12	-0·34	1·49
17	165·38	9·04	58·95	11·62	1·111	0·15	-0·806	1·20	0·950	0·10	-0·863	0·78
18	166·06	9·50	60·23	7·70	1·057	0·21	-1·217	1·63	0·950	0·10	-0·958	1·07

* 15 females and 15 males in each age group

BMD= bone mineral density, TBLH = total body less head

5.4.1 Diagnostic accuracy of MXA (420 iDXA)

Vertebral fracture assessment (VFA) of 5460 individual vertebrae was performed by R1 using the visual SQ method (this was the gold-standard read) and by R2 using the 33-point MXA technique; of these, 4% were not evaluable by either method because of poor image quality, including movement artefact. The majority of unevaluable vertebrae for both techniques were located in the upper thoracic spine (Figure 5.3).
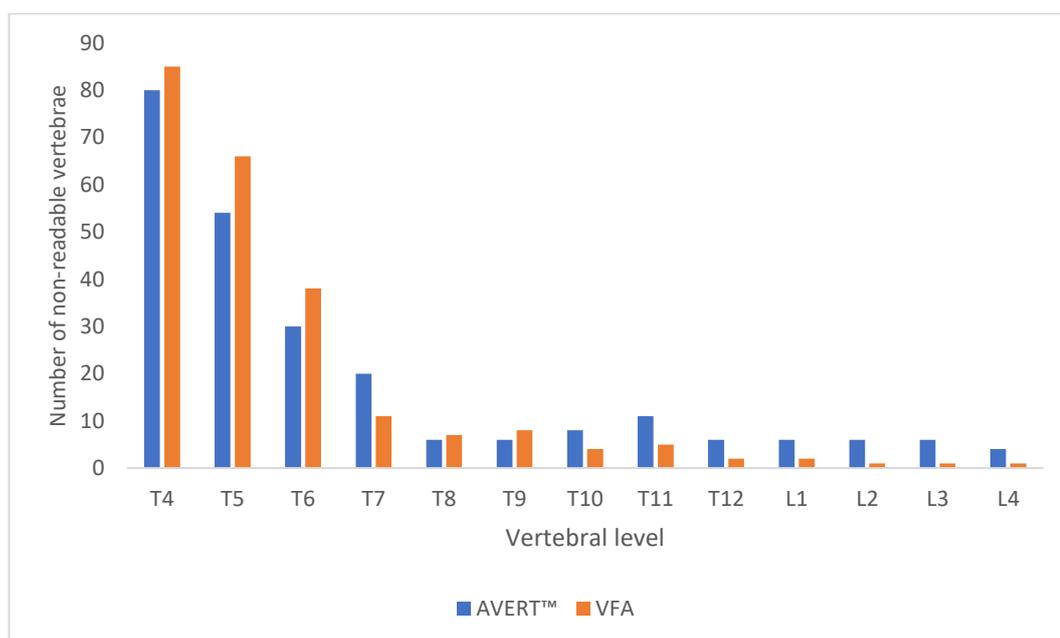
Figure 5.3 Total number of unevaluable vertebrae for VFA=231 (4%) and MXA (AVERT™) =243 (4%)

Among the 420 subjects, 191 (45%) had no fracture by the gold-standard visual SQ method, while mild, moderate and severe fractures were identified in 98 (23%), 67 (16%), and 29 (7%) subjects respectively. Isolated physiological wedging (with no fracture) was identified in 35 (8%) children. MXA identified more children with mild and moderate vertebral fractures than the gold standard but almost the same number of severe vertebral fractures. Table 5.2 shows the number and grading of the evaluated vertebrae by the two techniques.

Table 5.2 Prevalence of Vertebral Fractures in Study Cohort (n= 420 patients, 5460 vertebrae)

	VFA (R1) = gold standard		MXA– AVERT™ (R2)	
	Per vertebra	*Per subject*	*Per vertebra*	*Per subject*
No fracture	4564 (81%)	191 (45%)	4412 (78%)	157 (37%)
Mild fracture **(21% to 25% loss of height)**	216 (4%)	98 (23%)	441 (8%)	204 (49%)
Moderate fracture **(26% to 40% loss of height)**	124 (2%)	67 (16%)	317 (5%)	155 (37%)
Severe fracture **(≥ 41% loss of height)**	54 (1%)	29 (7%)	47 (1%)	27 (7%)
Non-readable vertebra	231 (4%)	80 (19%)	243 (4%)	98 (23%)
Fractures **(loss of height ≤ 20%)***	77 (1%)	32 (7%)	N/A	N/A
Physiological wedge	136 (3%)	35 (8%)	N/A	N/A
Possible fracture	58 (2%)	14 (14%)	N/A	N/A

* A height reduction of ≤ 20% that was nevertheless considered to represent a fracture rather than normal variation

The location of mild fractures and physiologically wedged vertebrae is shown in Figure 5.4.
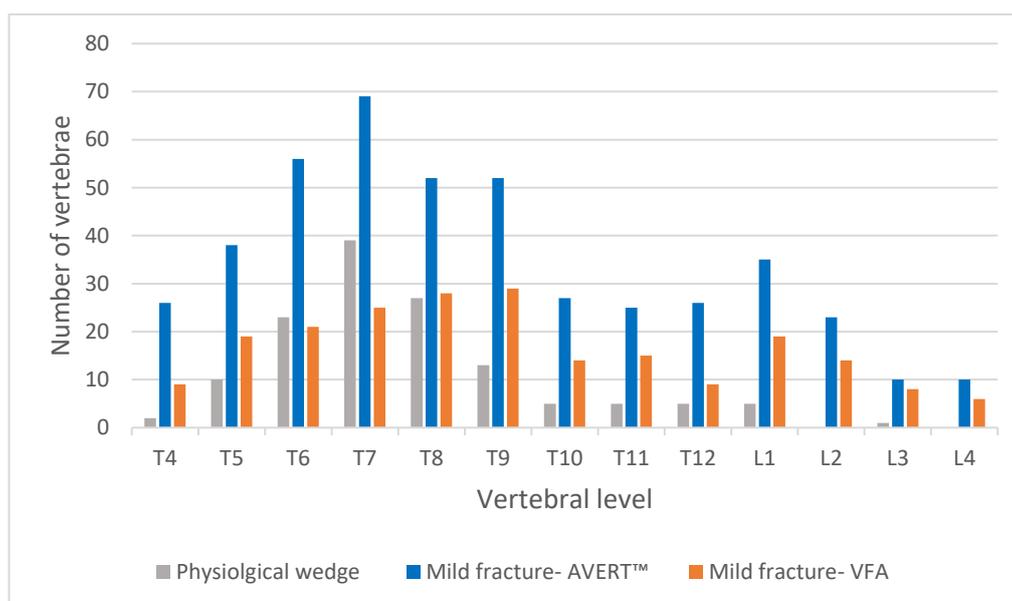
Figure 5.4 Number and location of mild vertebral fractures identified by both techniques compared to number of physiologically wedged vertebrae identified by VFA. The figure illustrates that in the mid-thoracic region, the number of mild fractures identified by AVERT™ was comparable to the sum of the mild fractures and physiological wedges identified by the visual SQ method (e.g. at T7 and T8, AVERT™ identified 69 and 52 mild fractures, respectively; whereas the sum of the mild fractures and physiologically wedged vertebrae identified by VFA were 64 and 54, respectively).

Figures 5.5a and b show the number, severity, and shape of vertebral fractures by the two methods at the vertebral and subject levels respectively, as well as the physiological wedges identified by VFA.
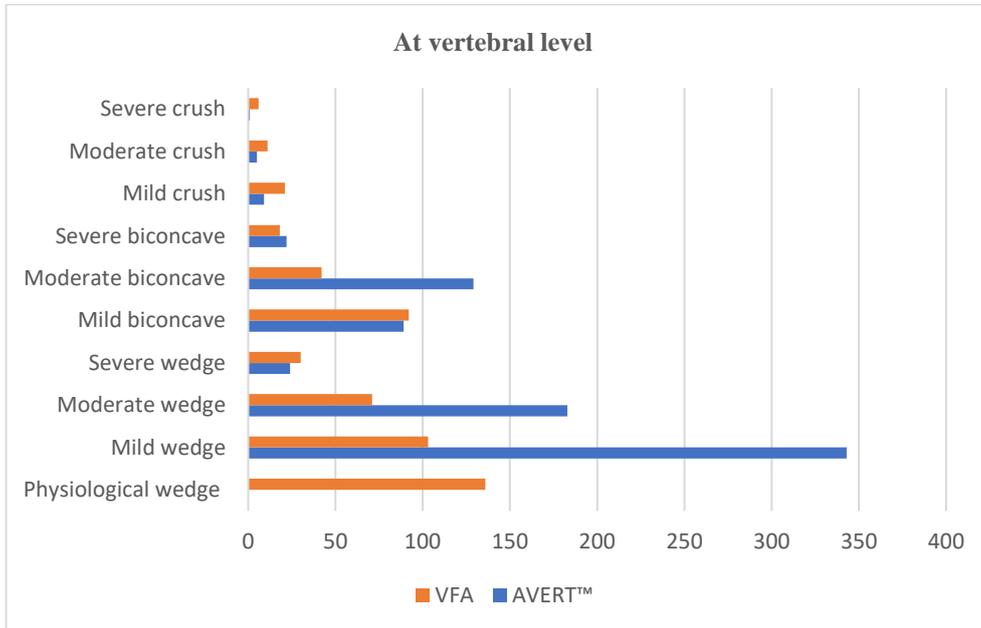
(5.5a)
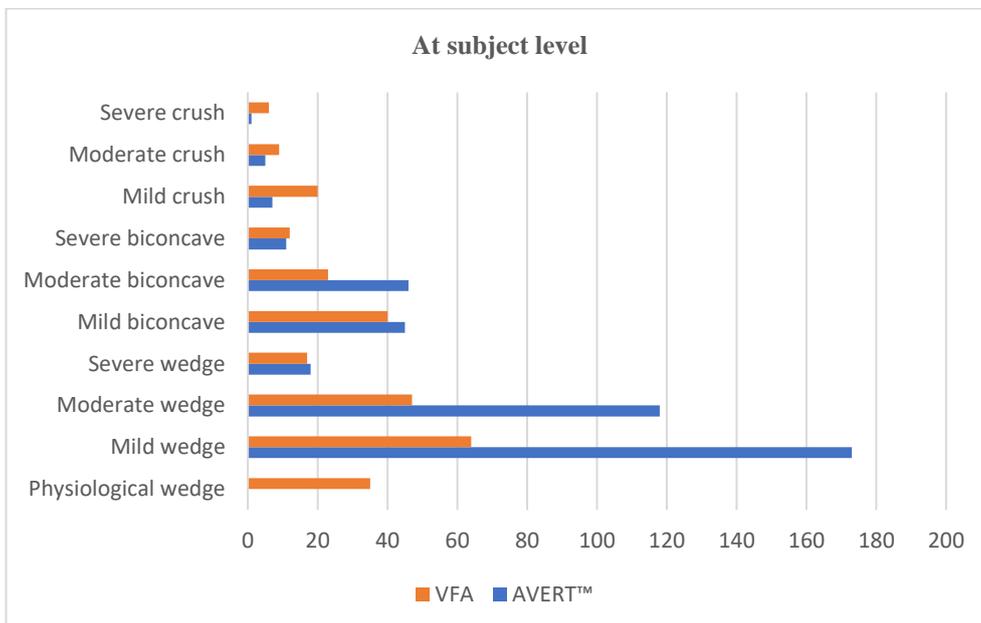
(5.5b)

Figure 5.5 Number of vertebral fracture shapes identified using both techniques (a) at vertebral level and (b) at subject level (note that AVERT™ does not have the ability to diagnose physiologically wedged vertebrae

The diagnostic accuracy and observer agreement of AVERT™ for the "any fracture" (≥ 21% loss of height), "moderate and severe fracture" (≥26% loss of height), and "mild fracture" (21% to 25% loss of height) groups are presented in Table 5.3.

Table 5.3 Diagnostic accuracy of MXA for detecting vertebral fractures (n = 420 patients, 5460 vertebra)

| Vertebral Level | Number of iDXA Scans Evaluable per vertebral level | Any fracture | | | | | moderate and severe fracture * | | | | | Mild fracture** | | | | |
		Sensitivity	Specificity	Agreement	Kappa	95% CI	Sensitivity	Specificity	Agreement	Kappa	95% CI	Sensitivity	Specificity	Agreement	Kappa	95% CI
T4	332	16/19 (84%)	289/313 (92%)	94%	0·50	0·33 – 0·66	4/4 (100%)	318/328 (97%)	94%	0·43	0·12 – 0·69	7/9 (79%)	304/323 (94%)	88%	0·37	0·21 – 0·55
T5	364	27/34 (79%)	286/330 (87%)	92%	0·44	0·31 – 0·57	8/10 (80%)	329/354 (93%)	95%	0·34	0·15 – 0·52	3/19 (16%)	310/345 (90%)	62%	0·03	-0·17 – 0·28
T6	388	39/45 (87%)	289/343 (84%)	93%	0·48	0·37 – 0·58	15/19 (79%)	347/369 (94%)	98%	0·49	0·30 – 0·65	8/21 (38%)	320/369 (87%)	82%	0·14	0·01 – 0·29
T7	400	46/52 (88%)	273/348 (78%)	94%	0·42	0·32 – 0·52	18/23 (78%)	343/377 (90%)	96%	0·42	0·28 – 0·55	11/25 (44%)	319/375 (85%)	82%	0·15	0·03 – 0·33
T8	404	45/52 (86%)	290/352 (82%)	96%	0·46	0·36 – 0·56	19/20 (95%)	325/384 (84%)	98%	0·32	0·20 – 0·44	7/28 (25%)	336/376 (89%)	66%	0·09	-0·14 – 0·21
T9	411	32/52 (64%)	317/361 (88%)	95%	0·41	0·29 – 0·52	12/17 (71%)	383/394 (97%)	95%	0·53	0·32 – 0·69	11/29 (38%)	344/382 (89%)	84%	0·20	0·09 – 0·33
T10	409	21/29 (72%)	347/380 (91%)	95%	0·45	0·30 – 0·58	10/11 (91%)	381/398 (96%)	95%	0·49	0·28 – 0·67	8/14 (57%)	377/395 (95%)	92%	0·36	0·21 – 0·44
T11	407	24/27 (89%)	348/380 (92%)	96%	0·53	0·39 – 0·65	10/10 (100%)	376/397 (95%)	96%	0·45	0·24 – 0·62	7/15 (47%)	375/392 (95%)	89%	0·31	0·17 – 0·48
T12	412	23/25 (92%)	357/387 (92%)	97%	0·54	0·40 – 0·67	13/14 (93%)	385/398 (97%)	97%	0·61	0·42 – 0·77	6/9 (67%)	384/403 (95%)	79%	0·32	0·19 – 0·43
L1	412	39/42 (93%)	335/370 (90%)	97%	0·63	0·52 – 0·72	18/19 (95%)	372/393 (95%)	97%	0·58	0·41 – 0·71	11/20 (55%)	369/392 (94%)	81%	0·36	0·22 – 0·49
L2	414	25/30 (83%)	361/384 (94%)	98%	0·60	0·45 – 0·72	13/14 (93%)	388/400 (97%)	97%	0·65	0·43 – 0·80	6/14 (43%)	383/400 (96%)	82%	0·29	0·13 – 0·36
L3	413	15/22 (68%)	384/391 (98%)	97%	0·66	0·46 – 0·82	8/10 (80%)	399/403 (99%)	98%	0·72	0·45 – 0·89	3/8 (38%)	398/405 (98%)	88%	0·31	0·19 – 0·48
L4	415	10/17 (59%)	389/398 (98%)	96%	0·53	0·30 – 0·70	5/7 (71%)	404/408 (99%)	96%	0·61	0·23 – 0·87	3/6 (50%)	402/409 (98%)	93%	0·36	0·22 – 0·49
Average		80%	90%	95%			87%	95%	96%			46%	92%	82%		
Subject level	420	131/133 (98%)	142/287 (49%)	92%	0·37	0·30 – 0·43	72/78 (92%)	227/342 (66%)	90%	0·43	0·35 – 0·51	93/96 (97%)	198/324 (61%)	79%	0·41	0·27 – 0·55

* moderate and severe (≥26%) vertebral height reduction; ** Mild fracture = 21% to 25% vertebral height reduction

5.4.2 Intraobserver agreement of MXA (100 iDXA)

There was fair to excellent intraobserver agreement, with kappa ranging from 0·49 to 0·87 (95% CI: 0·37, 0·98), with the lowest agreement level identified at T4. Figure 5.6 summarises intraobserver agreement of MXA for R2.
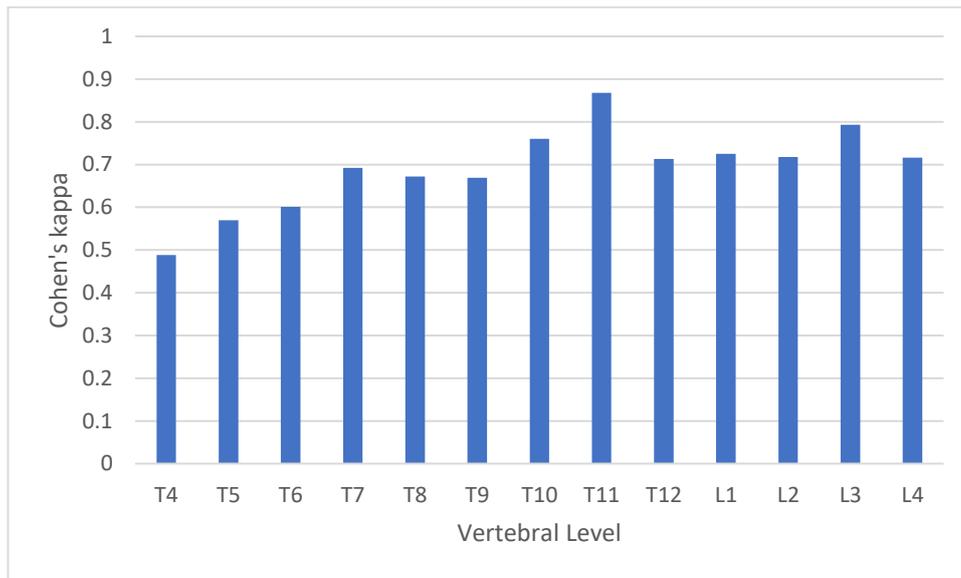
Figure 5.6 Intraobserver (R2) agreement of MXA/AVERT™

5.4.3 Observer agreement of MXA (30 iDXA)

In respect to the "any fracture" grade, there was fair to good interobserver agreement between the additional four raters when they used AVERT™, with kappa ranging from 0·39 to 0·53 (95% CI: 0·17 – 0·67). In contrast, there was a slightly higher agreement level when only "moderate and severe fractures" were considered, with kappa ranging from 0·48 to 0·67 (95% CI: 0·33 – 0·78). Finally, there was poor agreement when only "mild fractures" were considered, with kappa ranging from 0·10 to 0·29 (95% CI: -0·09 – 0·41). Intraobserver agreement for the same four readers for "any fracture" ranged from moderate to good, with mean kappa values for R1, R3, R4, and R5 of 0·55, 0·60, 0·68, and 0·58, respectively; for "moderate and severe fractures", kappa values

were 0·59, 0·82, 0·89, and 0·67 and for "mild fractures" kappa values were 0·67, 0·61, 0·51, and 0·58 respectively. Table 5.4 summarises inter- and intraobserver agreement of MXA for the four observers.

Table 5.4 Summary of inter and intraobserver agreement for MXA (n=30)

Interobserver agreement	Observer	Kappa		
Any fracture (≥ 21% loss of height)		Mean	Min	Max
	R1 vs R3	0·39	0·20	0·69
	R1 vs R4	0·44	0·23	0·57
	R1 vs R5	0·53	0·38	0·73
	R3 vs R4	0·41	0·20	0·85
	R3 vs R5	0·39	0·15	0·58
	R4 vs R5	0·42	0·11	0·70
	Agreement across four observers	Fleiss' kappa = **0·44**		
Moderate and severe fracture (≥ 26% loss of height)				
	R1 vs R3	0·50	0·30	0·76
	R1 vs R4	0·52	0·24	0·73
	R1 vs R5	0·67	0·42	0·92
	R3 vs R4	0·48	0·26	0·79
	R3 vs R5	0·56	0·36	0·96
	R4 vs R5	0·49	0·16	0·81
	Agreement across four observers	Fleiss' kappa = **0·52**		
Mild fracture (21% to 25% loss of height)				
	R1 vs R3	0·21	0·07	0·45
	R1 vs R4	0·21	0·01	0·43
	R1 vs R5	0·29	0·04	0·61
	R3 vs R4	0·19	0·04	0·56
	R3 vs R5	0·10	0·07	0·27
	R4 vs R5	0·15	0·03	0·51
	Agreement across four observers	Fleiss' kappa = **0·21**		
Intraobserver agreement	**Observer**	**Kappa**		
Any fracture (≥ 21% loss of height)				
		Mean	Min	Max
	R1	0·55	0·16	1·00
	R3	0·60	0·28	1·00
	R4	0·68	0·11	1·00
	R5	0·58	0·13	1·00
	Agreement across four observers	**0·60**	**0·11**	**1·00**
Moderate and severe fracture (≥ 26% loss of height)				
	R1	0·59	0·19	1·00
	R3	0·82	0·44	1·00
	R4	0·89	0·56	1·00
	R5	0·67	0·18	1·00
	Agreement across four observers	**0·74**	**0·18**	**1·00**
Mild fracture (21% to 25% loss of height)				
	R1	0·67	0·21	1·00
	R3	0·61	0·11	1·00
	R4	0·51	0·01	1·00
	R5	0·58	0·01	1·00
	Agreement across four observers	**0·59**	**0·01**	**1·00**

The average sensitivity, specificity, FP, and FN rates for the four observers were 89%, 79%, 21%, and 11% at the vertebral and 98%, 52%, 48%, and 2% at the subject level for any fracture grade. When only mild fractures were considered, the average sensitivity, specificity, FP, and FN rates were 36%, 86%, 14%, and 64% at the vertebral and 88%, 35%, 65%, and 12% at the subject levels respectively.

5.4.4 Observer agreement of VFA (100 iDXA)

Of the possible total of 1300 vertebrae, from T4 to L4 (i.e. 13 vertebrae per subject in 100 subjects); 1267 (97%) were adequately visualised by R1, and 1269 (98%), and 1248 (96%) by R2 and R3 respectively. The number and severity of vertebral fractures at the vertebral and subject levels for each observer are shown in Table 5.5. Although the numbers of mild and moderate vertebral fractures varied between all observers, the number of severe fractures was comparable. A similar pattern was observed at the subject level. Figure 5.7 summarises the interobserver agreement of VFA between R1 and R3.

Table 5.5 Fracture prevalence by observer and technique for 100 randomly selected images

	VFA (R1) = Gold standard		MXA– AVERT™ (R2)		VFA (R3)	
	Per vertebra	*Per subject*	*Per vertebra*	*Per subject*	*Per vertebra*	*Per subject*
No fracture	822 (63%)	32 (32%)	902 (69%)	11 (11%)	782 (60%)	14 (14%)
Mild fracture (21% to 25% loss of height)	149 (11%)	56 (56%)	176 (14%)	70 (70%)	208 (16%)	72 (72%)
Moderate fracture (26% to 40% loss of height)	97 (7%)	35 (35%)	153 (11%)	61 (61%)	130 (10%)	45 (45%)
Severe fracture ($\geq 41\%$ loss of height)	39 (3%)	19 (19%)	38 (3%)	20 (20%)	62 (4%)	22 (22%)
Non-readable vertebrae	66 (5%)	16 (16%)	31 (2%)	19 (19%)	70 (5%)	15 (15%)
Fractures (with loss of height $\leq 20\%$)*	55 (4%)	25 (2%)	N/A	N/A	48 (3%)	23 (2%)

* A height reduction of ≤ 20% that was nevertheless considered to represent a fracture rather than normal variation
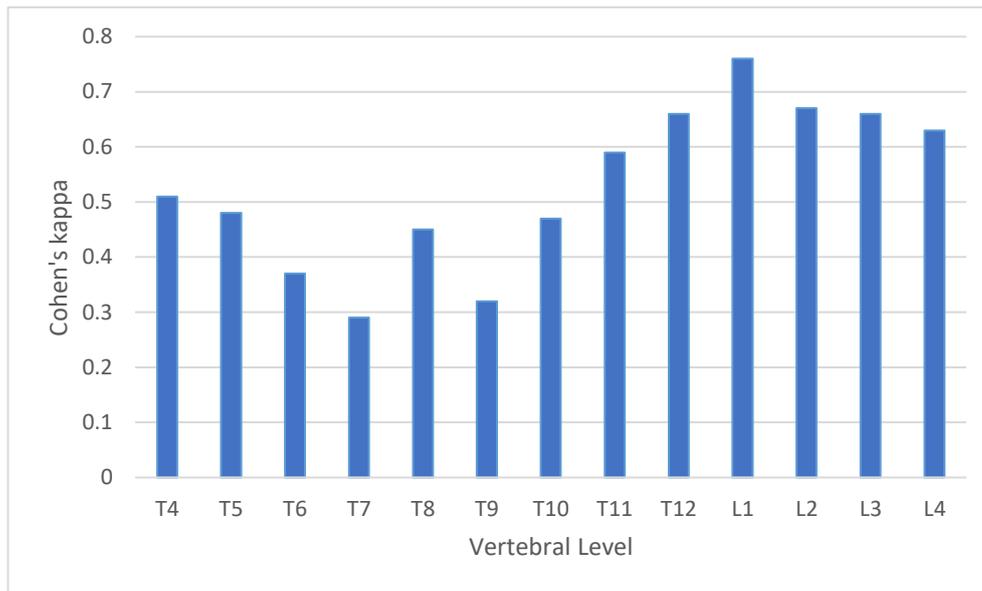
Figure 5.7 Interobserver (R1, R3) agreement of VFA

5.5 Discussion

This study aimed to determine the diagnostic accuracy and inter and intraobserver agreement of morphometric vertebral fracture analysis (MXA) using a 33-point software programme (designed for adults) on a large cohort of children with conditions predisposing to vertebral fracture. Results of MXA were compared to the visual SQ technique for vertebral fracture identification from iDXA scans (VFA). Results demonstrate that MXA is only as good as VFA in identifying severe vertebral fractures with reduced diagnostic accuracy for detecting mild vertebral fractures.

The overall sensitivity, specificity, FP, and FN rates for R2 to R5 were 89%, 79%, 21%, and 11% at the vertebral and 98%, 52%, 48%, and 2% at the subject levels. Five previous studies that used 6-point MXA and VFA [5–8, 11] have shown generally lower diagnostic accuracy, except for a higher specificity at subject level than has been shown by the current study. This may be due to the high number of subjects with physiological wedging according to the reference standard which were diagnosed as mild fractures by MXA, thus causing an increase in the false positive rate for MXA.

151

The results of observer agreement of MXA in this current study are slightly higher than those of a previous study, for which the evaluation was conducted by three readers (an experienced clinical scientist, a senior radiographer and a clinical scientist unfamiliar with MXA) [6]. In that study, kappa scores ranged from 0·13 to 0·32 when compared to VFA. On the other hand, our results show a slightly lower agreement level when compared to another recent study [8], where kappa reached 0·79 (95% CI: 0·62 – 0·92) and 0·55 (95% CI: 0·40 – 0·68) at the vertebral and subject levels, respectively. It should be noted that the study was based on only 20 subjects, and the gold standard was radiographic images reported by a non-radiologist reader [8]. Yet another study used Hologic QDR Physician's viewer software (version 7·02) to perform MXA on lateral DXA scans of 58 children and adolescents, using six-point software. This reported higher agreement at both the vertebral and subject levels (a kappa score of 0·72 (95 % CI: 0·65 – 0·78), and 0·73 (95 % CI: 0·55 – 0·91) respectively) when compared to the visual SQ method using conventional radiographs and performed by two experienced skeletal radiologists [7]. Notably, no comparison was established in that study between MXA and visual SQ for VFA. Finally, our current findings are better than those of a recent study on radiographic images of 137 children, in which five observers utilised a six-point software programme (SpineAnalyzer™, Optasia Medical, Cheadle, UK); kappa for interobserver reliability ranged from 0·05 to 0·47 (95% CI: -0·19 – 0·76) and the intraclass correlation coefficient for intraobserver reliability ranged from 0·25 to 0·61[11].

Despite improvement in diagnostic accuracy of 33-point MXA compared to 6-point MXA and VFA, our results show low diagnostic accuracy and observer reliability when only mild fractures are considered. Our results suggest that a large contributory factor is the inability of the software to distinguish normal physiological wedging (i.e.

developmental morphological variability that occurs throughout childhood) from mild fractures, particularly in the thoracic region. As a consequence, the rate of mild and moderate fracture was relatively higher for MXA than for the reference standard.

This inability to differentiate normal physiological wedging from fracture contributes to the generally low observer reliability of vertebral fracture diagnosis in children, irrespective of imaging modality or scoring system [5]. Software that is developed on a healthy cohort of children which incorporates relevant variables related to age may be the solution to accurate and reliable diagnosis of mild vertebral fractures in children.

Another major limitation of MXA is the inability of the software programme to identify fractures when height loss is below 20%, identified in 32 subjects (8%) in this study.

It should be pointed out that observer reliability of MXA depends on point placement, which to a large extent affects thresholds for height ratios. In other words, only a very small alteration in point placement and therefore in height ratio (that would be insignificant clinically) can lead to two different fracture categories being reported by 2 observers or by the same observer at different times (e.g. 24·9% and 25·1% loss of height will be classified as mild and moderate fractures respectively). This is particularly important at the threshold between no vertebral fracture and mild vertebral fracture.

The T5-T9 region evidenced the most fractured levels on the spine. This is in line with previous research on children that has reported that the mid and upper thoracic spine are the most fractured levels in children [6,7,10,11]. However, in practice, the exact location to place points in the thoracic region can be problematic as the region significantly affects the level of noise, mainly due to soft tissue and bony structures

that may affect the operators' visualisation. Therefore, this might have increased the rate of false positives.

Considering individual vertebral levels, the L1–L4 region showed the highest kappa scores, indicating that the lower vertebral levels are more adequately visualised and more likely to be assessed correctly by all observers using the two methods. This is in line with previous research that has reported on the difficulty of identifying vertebral fractures in the mid and upper thoracic spine in children [8, 9, 11].

A limitation of this study is that the rating of only one experienced paediatric radiologist was used as reference rather than a consensus of several radiologists. However, only a single radiologist provides the clinical report, so in this respect the study design more closely resembles clinical practice. The subjectivity of positioning the points on each vertebral body is a limitation of any quantitative morphometric technique and cannot easily be avoided. This is further complicated in children who have age-dependent changes in vertebral body ossification. A clear guideline as to where the points should be positioned in children prior to full vertebral ossification is required. The strength of this current study is that it demonstrates the utility of the 33-point software programme to conduct MXA in the hands of various observers, including three paediatric radiologists, a radiographer and a clinical scientist, all with varying degrees of experience. With a reliable software programme, specifically designed for use in children, non-medical staff could be trained to perform MXA. However, as emphasised by a previous study [15], a second read by a radiologist is required to reliably differentiate mild fractures from non-fracture deformities. Endplate changes and ossification process can be problematic for vertebral fracture identification in children; therefore, the authors believe that visual methods such as the ABQ approach are more accurate methods of assessing vertebral fractures in children.

5.6 Conclusion

MXA reaches only moderate agreement when compared to the visual SQ VFA technique, with fair to moderate inter and intraobserver agreement. In order to facilitate the detection of mild vertebral fractures in children, a paediatric standard is required which not only incorporates specific vertebral body height ratios but also the age-related physiological changes in vertebral shape that occur throughout childhood.

5.7 Acknowledgments

5.8 References

1. Bishop N. Characterising and treating osteogenesis imperfecta (2010) Early Hum Dev 86:743-6

2. Huber AM, Gaboury I, Cabral DA, et al (2010) Prevalent vertebral fractures among children initiating glucocorticoid therapy for the treatment of rheumatic disorders. Arthritis Care Res 62:516-26

3. Halton J, Gaboury I, Grant R, et al (2009) Advanced vertebral fracture among newly diagnosed children with acute lymphoblastic leukemia: Results of the Canadian steroid-associated osteoporosis in the pediatric population (STOPP) research program. J Bone Miner Res 24:1326-34

4. Bishop N, Arundel P, Clark E, et al (2014) Fracture Prediction and the Definition of Osteoporosis in Children and Adolescents: The ISCD 2013 Pediatric Official Positions. J Clin Densitom 17:275–80

5. Adiotomre E, Summers L, Allison A, et al (2017) Diagnostic accuracy of DXA compared to conventional spine radiographs for the detection of vertebral fractures in children. Eur Radiol 27:2188-99

6. Crabtree N, Chapman S, Högler W, et al (2017) Vertebral fractures assessment in children: evaluation of DXA imaging versus conventional spine radiography. Bone 97:168–74

7. Diacinti D, Pisani D, D'Avanzo M, et al (2015) Reliability of vertebral Fractures assessment in children with osteogenesis imperfecta. Calcif Tissue Int 96:307-12

8. Kyriakou A, Shepherd S, Mason A, Ahmed SF. (2015) A critical appraisal of vertebral fracture assessment in paediatrics. Bone 81:255-9

9. Siminoski K, Lentle B, Matzinger MA, Shenouda N, Ward LM, Canadian SC (2014) Observer agreement in pediatric semiquantitative vertebral fracture diagnosis. Pediatr Radiol 44:457-66

10. Adiotomre E, Summers L, Allison A, et al (2016) Diagnosis of vertebral fractures in children: is a simplified algorithm-based qualitative technique reliable? Pediatr Radiol 46:680-8

11. Alqahtani FF, Messina F, Kruger E, et al (2017) Evaluation of a semi-automated software program for the identification of vertebral fractures in children. Clin Radiol 72:e904– 11

12. Bromiley PA, Adams JE, Cootes TF (2015) Automatic localisation of vertebrae in DXA images using random forest regression voting. In: Vrtovec T. et al. (eds) International Workshop on Computational Methods and Clinical Applications for Spine Imaging; 2015: Springer. Cham

13. Lindner C, Bromiley PA, Ionita MC, Cootes TF (2015) Robust and accurate shape model matching using random forest regression-voting. IEEE Trans Pattern Anal Mach Intell 37:1862-74

14. Genant HK, Wu CY, Vankuijk C, Nevitt MC (1993) Vertebral fracture assessment using a semiquantitative technique. J Bone Miner Res 8:1137–48

15. Zeytinoglu M, Jain RK, Vokes TJ (2017) Vertebral fracture assessment: Enhancing the diagnosis, prevention, and treatment of osteoporosis. Bone 104:54-65

Chapter Six
Overall summary and discussion

6.1 Overall summary and discussion

Patients with reduced bone density are more at risk of experiencing fractures, even after minor traumatic events. The most common types of fractures in these cases are in the vertebrae and femoral necks in adults, and vertebrae in children. Therefore, the key issue which has prompted the research presented in this thesis is that the dependence of the diagnosis of osteoporosis in children on the identification of vertebral fractures and therefore it is important that vertebral fractures are promptly and reliably diagnosed. Vertebral fractures are different from osteoporotic fractures of the limbs in the sense that they are normally silent. Added to this is that if they are not treated, progressive degeneration of body height occurs. However, if diagnosed early, medical care can be initiated immediately in the form of treatment with bisphosphonates that limit the development of incident fractures.

Thus, one of the aims of this thesis was to report and review the results of previous research studies in relation to the diagnosis of osteoporotic vertebral fractures in children. However, there is no accepted standardised technique for detecting vertebral fractures in children. Commercially available semi-automated programmes have been trialled on adults with previous studies demonstrating that such programmes are reliable and accurate in measuring vertebral height and identifying vertebral fractures in adults (Chapter 4, page 124). As there is yet no specific semi-automated software for use with children, the general aim of this thesis was to evaluate the available semi-automated software programmes that exist for adults and to compare the results with the gold standard method currently used in clinical practice (visual assessment by radiologists).

In terms of the results related to inter and intraobserver agreement, the first finding from this thesis confirms the findings from existing literature showing significant inter

and intraobserver variability in diagnosing vertebral fractures in children. This was confirmed using the SpineAnalyzer™ software programme. In Chapter Three (page 89), the results indicate a relatively low inter and intraobserver agreement, with kappa ranging from 0.05 to 0.47 (95% CI: -0.19, 0.76) and ICC ranging from 0.25 to 0.61. The overall sensitivity and specificity were 18% (95% CI: 14–22) and 97% (95% CI: 97–98), respectively. Reasons behind this low level of agreement may be the subjectivity of point placement by different observers. Here, experience was certainly a determining factor in the differences when identifying T12/L1 (the lowermost vertebral body associated with ribs was identified as T12). Another reason may be the method used by SpineAnalyzer™ in identifying vertebral fractures. This is based only on the loss of height of vertebral bodies whilst the gold standard is sABQ, a visual method that considers alterations in the vertebral endplates that may be non-fracture related. On the basis of these results, the present author recommends the placing of more than six points (the software limit of SpineAnalyzer™) to accurately represent vertebral morphometry in children.

In Chapters Four and Five, AVERT™, which is a new 33-point technique semi-automated software programme, was used on paediatric radiographs and DXA-VFA. AVERT™ is made available through a collaboration between Optasia Medical, a medical image analysis company who developed and distributed the earlier SpineAnalyzer™ package and the University of Manchester (UoM). Currently, the software is in use at the UoM (for adults) and the University of Sheffield (UoS) (for children) for development purposes. Ultimately, all parties aim to develop a fully automated computer system for identifying vertebral fractures (in both adults and children). Optasia Medical provided the main author with a free license for AVERT™ for training purposes.

The aim of Chapter Four (page 108) was to evaluate whether observer reliability and diagnostic accuracy of MXA for the identification of vertebral fractures in children is improved with a 33-point semi-automated programme (AVERT™) compared to the 6-point programme (SpineAnalyzer™) utilising both VFA and radiographs on the same 50 subjects from the previous experiments (Chapter Three, page 81). Overall, poor to fair agreement across the four techniques was found when compared with the reference standard in terms of identifying vertebral fractures: the average kappa score ranged from 0.26 to 0.37 (95% CI: 0.16, 0.46), with XR AVERT™ having the highest value followed by DXA AVERT™.

The overall sensitivity of SpineAnalyzer™ and AVERT™ ranged from 26% to 31% and 36% to 41% respectively, and the overall specificity ranged from 96% to 98% and 91% to 95% respectively. Comparing these findings to those of the previous chapter (Chapter Three, page 89), AVERT™ has a slightly increased diagnostic accuracy compared to SpineAnalyzer™ for both DXA and XR; however, both programmes showed low diagnostic accuracy, poor technique agreement and high false negative rates. This study further confirms findings from previous studies indicating significant inter and intraobserver variability in diagnosing vertebral fractures in children.

For Chapter Five (page 133), the plan was to assess and train AVERT™ using a large database of iDXA scans (420 DXA scans) enriched with a high fracture prevalence. This allowed evaluation of the MXA method under realistic operating conditions, such as would be encountered in clinical use in the hands of various observers. The reference standard was a visual assessment relying on Genant's SQ method on VFA by an experienced paediatric radiologist. When compared to the gold standard, the overall sensitivity, specificity, FP, FN rates and the degree of agreement (reported by kappa) were 80%, 90%, 10%, 20% and 0.37 respectively for any fracture (mild, moderate and

severe) group. However, a slight improvement in all diagnostic accuracy parameters was seen when calculated for "moderate and severe fracture" group; 87%, 95%, 5%, 13% and 0.43 respectively. In contrast, for the mild fracture group, values of 46%, 92%, 8%, 54% and 0.41, respectively were obtained. Additionally, a clinical scientist reported 100 VFA by a visual SQ grading: the interobserver agreement between two expert readers ranged from fair to good [kappa = 0.29 to 0.76 (95% CI: 0.17 – 0.88), with T7 and T9 scoring the lowest kappa values of 0.29 and 0.32, respectively. The inability to differentiate normal physiological wedging from fracture also accounts for low observer agreement of VFA. Finally, observer reliability of AVERT™ was evaluated more widely by four observers (three paediatric radiologists and an experienced clinical scientist) on 30 DXA-VFA. For the "moderate and severe fracture" group, inter and intraobserver agreement across the four observers ranged from 0.48 to 0.67 (95% CI: 0.37, 0.75) and from 0.59 to 0.89 (95% CI: 0.49, 0.97), respectively. However, when only mild fractures were considered, the inter and intraobserver agreement across the four observers ranged from 0.10 to 0.29 (95% CI: 0.01, 0.37) and from 0.51 to 0.67 (95% CI: 0.42, 0.76), respectively. Despite improvement in diagnostic accuracy of 33-point MXA compared to 6-point MXA, our results showed low diagnostic accuracy and observer reliability when only mild fractures were considered. Severe fractures may lead to spinal deformity, so it is important to identify and treat when they are mild.

From the results of all research chapters presented in this thesis, kappa values for inter and intraobserver agreement and diagnostic accuracy were significantly higher for the lower vertebral levels (L1-L4) than for the mid and upper thoracic levels. These findings support those of previous studies reporting that visualisation and identification of vertebral fractures in the mid and upper thoracic spine is one of the major challenges

for fracture diagnosis in children. The reasons for this difficulty have been fully discussed in Chapter Three (page 98).

These studies were made possible due to the cooperation between different parties including; UoS, UoM, SCH, Birmingham Children's Hospital NHS Foundation Trust and Optasia Medical Limited. The outcomes reported in this thesis show the need to enhance reliability of semi-automated software programmes for use in children. It is this author's opinion that these findings have established a strong foundation on which to build future SARACEN studies (see 6.3, page 165 for more details).

In conclusion, the findings reported in this thesis have demonstrated low inter- and intraobserver reliability and diagnostic accuracy for vertebral fracture identification in children for semi-automated software programmes that exist for use with adults. Neither AVERT™ nor SpineAnalyzer™ (existing adult software programmes) are satisfactorily reliable for vertebral fracture diagnosis in children. Although the programmes may appear useful in adults, the systems need revision by being trained with paediatric images. A specific algorithm should be designed to determine point placement and incorporate overall vertebral body shape, and the classification needs to be based on a grading system specifically designed to differentiate physiological variation from mild vertebral fractures. Development of specific paediatric software and normative values (incorporating age-related physiological variation in children) is therefore required.

6.2 Challenges and limitations

The most significant challenge of this thesis is the retrospective nature of all studies. Study samples for Chapters Three and Four (pages 76 and 106) were limited to the availability of images with the reference standard results established from a previous

study (three paediatric radiologists' readings). For the results presented in Chapter Five (page 142) however, there were enough subjects to include older children and adolescents (5 – 18 years old) for males and females and thus a new reference standard was established by an experienced paediatric radiologist. Another challenge was related to differences in identifying T12/L1 for the observers in Chapters Three and Four. This issue may increase the variability in observer reliability. In Chapter Five however, this limitation was countered by having a marker placed adjacent to an agreed vertebra (an arrow added to demonstrate the T12 vertebral body, lowest vertebral body associated with a rib) so that all observers recognised the same vertebral levels.

Another challenge that could not easily be avoided was the poor visualisation of the upper vertebral levels, which caused difficulty for the observers to apply the programmes and contributed to variability in point placement. As both programmes (AVERT™ and SpineAnalyzer™) only accept Digital Imaging and Communications in Medicine (DICOM) images, an additional challenge was the use of an external programme (Horos) to convert the format of lateral spine DXA collected from the PACS at the Radiology Department of SCH, from Joint Photographic Experts Group (Jpeg) to DICOM. This step may have contributed to reducing the spatial resolution of images.

The main limitation of this PhD study was the lack of an objective gold standard to compare with the results of semi-automated software programmes for diagnosing vertebral fractures in children. Therefore, the authors were not sure if the vertebrae were truly fractured in the case of (potential) mild fractures. The best solution available to us was using reads of experienced paediatric radiologists as reference standard (clinical daily practice method). Another limitation of this study, was the subjectivity in placing the landmark points on the vertebral body. It should be noted that this is a

weakness of any quantitative morphometric technique. Despite the training provided to observers before commencing the study, identifying vertebral morphometry points often varies between observers depending on their knowledge, experience and skills. The issue with this subjectivity is that small differences in positioning the points may affect the overall results. This point has been discussed with the inclusion of illustrative figures in Chapter Three (pages 93-94).

Finally, while the UoS and SCH team was able to establish a reference standard for 420 DXA and to evaluate AVERT™ using paediatric images, there was an issue securing sufficient funds to continue the study and to further develop such software. The Computing Department at the UoM was not able to secure funds to carry on the research training of AVERT™ using the paediatric images and thus returned the untrained software to our team for further testing of the programme. This point is discussed in more detail in the next section.

6.3 Future work

There are numerous avenues for future research which readily suggest themselves from the results of this thesis. I believe continuation of this work using the following guidelines will provide the opportunity for the betterment of our understanding of diagnosis of vertebral fractures in children. Firstly, in terms of future work that could be carried out with the Computing Department at UoM, all data has been transferred to the Manchester team. Specific software will be used to extract all data including point locations on each vertebral body and clinical diagnosis (reference standard) from AVERT™ and convert them into text files that can be loaded into other software for analysis and model building. The aim would be to use the manual annotations of points on the vertebrae (13 vertebrae, 33 points on each vertebral body leading to 429 points

for each subject) to build RFRV-CLM models and test their accuracy (in terms of the accuracy of automatically localising points on query images). Then, after training the new model, the software could be run again on the same 50 images (Chapter Four) that were tested before the training process to see if the results improve. Finally, once the tool is developed, conducting a cohort study/prediction of vertebral fracture study would be the optimal approach for assessing the tool in children. Numerous additional experiments would also be possible, such as testing whether age-specific models lead to higher classification accuracy.

Another key avenue of exploration is suggested from our further demonstration of poor inter and intraobserver reliability of the existing adult software tools (SpineAnalyzer™ and AVERT™). In order to improve this aspect of diagnosis, we suggest improving the outline of the vertebral bodies as the first stage of the annotation process. Therefore, we plan a collaboration with Dr. Arul Selvan of the Faculty of Arts, Computing, Engineering and Sciences at Sheffield Hallam University to use a method known as hierarchical clustering-based segmentation (HCS) to more precisely outline the boundaries of individual vertebral bodies. We will provide Dr. Selvan with the same anonymised DXA scans of children used in previous phases of the study. However, the images will need to be downloaded and burnt to CD directly from the DXA scanner. This process will be performed by radiographers at SCH responsible for obtaining the initial DXA images. The author will provide the anonymisation codes to the radiographers so that images from this and previous phases of the study will be linked. Minor amendments to the ethical approval of the study have already been made by the Health Research Authority and accepted by the SCH Trust to allow this.

The absence of an objective reference standard for "normal" vertebrae in children is a significant issue. As far as we know, no study has objectively assessed variability of

vertebral shape in healthy children. Therefore, our team recommend the development of normative data for vertebral morphology in children by gathering healthy children's spine images (DXA-VFA) to study change in vertebral morphology with growth. This would allow more valid and accurate identification of mild vertebral fractures in children with long-term conditions causing decreased bone density, and who are hence at a greater risk of fracture. The reference standard should incorporate variables including; age, sex, maturity, body mass index, bone age and BMD, as these might affect vertebral morphometry. This will help to build 2-dimensional models from DXA scans. Then, 3-dimensional models can be built by using existing CT images of patients with normal vertebrae. However, all images should be reviewed by experts in order to achieve the most ideal results.

Through the adoption of these suggestions, the author believes that children at risk of vertebral fracture can then have their fractures identified more reliably because clinicians will be better able to identify abnormalities and initiate treatment.

Chapter Seven

Appendices

7.1 Appendix 1 – Abstracts, Posters and Awards

P-162   Evaluation of the SpineAnalyser software programme on radiographic images for children
Fawaz Alqahtani; Amaka Offiah
Department of Oncology and Metabolism, Academic Unit of Child Health, Sheffield Children's NHS Foundation

Purpose: There is significant inter and intraobserver variability in diagnosing vertebral fractures (VF) in children, with a need to develop more objective methods. Semi-automatic software programmes such as SpineAnalyser may be the solution.

Methods: VF diagnosis was performed independently by five observers using the SpineAnalyser software from T4 through to L4 from the lateral spine radiographs of 137 children and adolescents with a median age of 12 years (range 5-15). A previous consensus read by 3 paediatric radiologists using a simplified ABQ technique (i.e. no software involved) served as the reference standard.

Results: Of a total of 1781 vertebrae, 1187 (66.64%) were adequately visualised by 3 or more observers. T5 was the most unreadable level 37.22% (51/137) and the two highest visualised levels were L2 and L4 (82.48%, 113/137 and 81.02% 111/137 respectively). Diagnostic accuracy (sensitivity, specificity and 95% confidence intervals) and inter observer reliability (Cohen's kappa) calculations of SpineAnalyser are on-going.

Conclusion: There was relatively good readability of vertebral bodies of mid thoracic and lumbar spine. However visibility was somewhat limited in the upper thoracic spine. Reasons included the summation caused by intrathoracic tissues and shoulders; poor image quality; and patient positioning. Once data analysis has been completed, we will present sensitivity, specificity and observer reliability of a software tool compared to routine qualitative radiographic analysis for VF diagnosis in children.

Abstract 7.1.1a Abstract submitted and accepted for the United Kingdom Radiology Congress (UKRC), Liverpool, UK 2016

Evaluation of the SpineAnalyzer Software Programme on Paediatric Radiographic Images

Fawaz F Alqahtani[1], Fabrizio Messina[3], Elzene Kruger[2], Heerunpal Gill[4], Michael Ellis[4], Isla Lang[2], Penny Broadley[2], Amaka Offiah[1][2]

[1] Academic Unit of Child Health, University of Sheffield, [2] Radiology Department, Sheffield Children's NHS Foundation Trust, [3] School of Health and Related Research, University of Sheffield, [4] Sheffield Medical School, University of Sheffield, Sheffield, UK

Contact: Fawaz. ffmalqahtani1@Sheffield.ac.uk

Background

- A common osteoporotic fracture is the vertebral fracture (VF)
- The detection of one or multiple VF (identified by a 20% reduction in vertebral body height) is indicative of bone fragility, irrespective of the reported bone mineral density (BMD)[1]
- Unlike osteoporotic fractures of the limbs, VF are typically silent and if untreated will lead to progressive loss of vertebral body height
- However if VF are diagnosed early, bisphosphonate treatment can help to treat present and stop the occurrence of new VF
- There is significant inter and intraobserver variability in diagnosing VF in children, with a need to develop more objective methods
- Semi-automatic software programmes such as SpineAnalyzer may be the solution

Aim

To evaluate the clinical utility of SpineAnalyzer by investigating the readability and accuracy of identifying VFs in a group of children and adolescents undergoing routine clinical assessment

Method

- 137 existing radiographs were collected from children aged between 5 to 15 years recruited into a previous study[2]
- SpineAnalyzer was used by five observers to independently analyse lateral spine images as follows:
a. The reader identifies T12 (lowest visible ribs) then places a point in the centre of all vertebrae between T4 and L4 (Fig 1); the software numbers the vertebral bodies accordingly and assigns 6 points as shown (Fig 1a)
b. The vertebral body outline is generated by manually adjusting the 6 points to record the anterior, middle and posterior vertebral heights (Fig 1b)
c. The software generates a deformity results table (Fig 1c)
- A previous consensus read by 3 paediatric radiologists using a simplified ABQ technique (i.e. no software involved) served as the reference standard used to compare the results obtained by SpineAnalyzer[2]

- **Statistical Analyses**
- Kappa statistics and 95% confidence intervals (CI) calculated on evaluable vertebral bodies
- Sensitivity and specificity of the software tool
- R software was employed to conduct data analyses
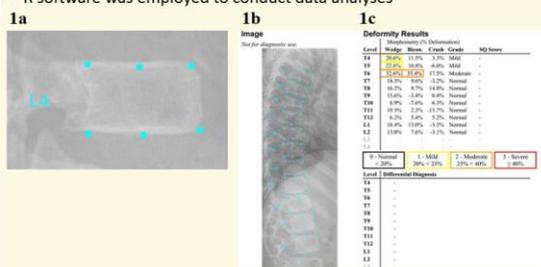
Figure 1: SpineAnalyzer analysis
The six manually identified points to outline the vertebral body (Figure 1a) are utilised to identify all vertebrae from T4 to L4 (Figure 1b), then to calculate values reported in the deformity results table (Figure 1c)

Results for SpineAnalyzer

- Mean age of the 137 subjects was 12.0 years (range, 5 to 15); 67% (92/137) were female
- Of a total of 1781 vertebrae, 1187 (67%) were adequately visualised by 3 or more observers (Fig 2)
- T4 was the most unreadable level 83/137 (61%) and the two highest visualised levels were L2 and L4 113/137 (82%), and 111/137 (81%) respectively
- Interobserver agreement in vertebral readability (average kappa score) for each vertebral level ranged from 0.054 to 0.476 (Fig 3)
- Overall sensitivity was 18% (95% CI 14 – 22) and overall specificity was 97% (95% CI 97 – 98) as shown in Table 1
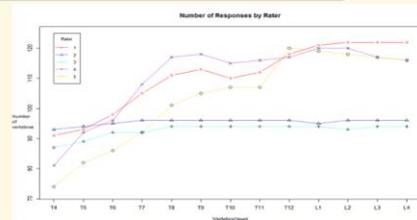
Fig. 2. The number of readable vertebrae for individual observers, showing a trend towards increased readability from T to L spine
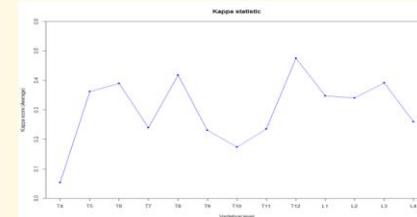
Fig.3. Average kappa score between the five observers using SpineAnalyzer (T4 achieved the lowest agreement)

	Sensitivity	95% CI	Specificity	95% CI
T4	18.20%	(13.22 - 23.18)	96.27%	(90.91 - 101.63)
T5	20.56%	(11.25 - 29.87)	98.75%	(96.33 - 101.17)
T6	**32.20%**	(16.70 - 47.70)	96.51%	(91.29 - 101.73)
T7	17.25%	(7.25 - 27.25)	96.06%	(92.51 - 99.60)
T8	22.80%	(12.21 - 33.3)8	95.70%	(92.56 - 98.84)
T9	18.22%	(6.39 - 30.04)	98.25%	(96.90 - 99.61)
T10	10.17%	(3.40 - 16.94)	98.10%	(96.43 - 99.77)
T11	19.50%	(9.11 - 29.89)	**95.62%**	(92.40 - 98.84)
T12	24.74%	(13.23 - 36.24)	96.04%	(92.98 - 99.10)
L1	17.3%	(9.97 - 24.67)	98.32%	(95.88 - 100.77)
L2	13.70%	(3.61 - 23.78)	98.59%	(96.57 - 100.61)
L3	**7.93%**	(2.41- 13.45)	99.06%	(97.87 - 100.26)
L4	10.30%	(3.41 - 17.20)	**99.32%**	(98.56 - 100.09)

Table 1. Sensitivity and specificity (95% CI) for T4 to L4

Figure 4a illustrates agreement. All 5 observers identified a severe fracture at T8

Figure 4b illustrates disagreement. T5, T6, and T7 were deemed non-evaluable by one observer, two observers called these mild, one observer moderate fracture while the fifth observer identified no fractures

Discussion

- There was relatively good readability of vertebral bodies of mid thoracic and lumbar spine but slightly limited in the upper part of the thoracic spine. Reasons include the summation caused by intrathoracic tissues and shoulders; poor image quality; and patient positioning
- SpineAnalyzer had an overall sensitivity of only 18% and specificity of 97%
- Low sensitivity may be due to
1. The subjectivity related to the semi-automated placement of six points, which may vary in position between readers
2. The use of the Genant system[3] to calibrate the software (the Genant system was developed from spine radiographs of osteoporotic women)

Conclusion

- SpineAnalyzer needs training on paediatric radiographs
- Observers need training on point placement
- The six-point approach may not be reliable enough for VF diagnosis in children

References

1. Bishop, N., Arundel, P., Clark, E, Dimitri, P., Farr, J, Jones, G., Makitie, O., Munns, C. F. & Shaw, N. 2014. Fracture Prediction and the Definition of Osteoporosis in Children and Adolescents: The ISCD 2013 Pediatric Official Positions. J Clin Densitom, 17, 275-280
2. Adiotomre, E., Summers, L., Allison, A., Walters, S. J., Digby, M., Broadley, P., Lang, I. & Offiah, A. C. 2016. Diagnosis of vertebral fractures in children: is a simplified algorithm-based qualitative technique reliable? Pediatr Radiol, 46, 680-688
3. Genant, H. K., Wu, C. Y., Vankuijk, C. & Nevitt, M. C. 1993. Vertebral fracture assessment using a semiquantitative technique. J Bone Miner Res, 8, 1137-1148

Poster 7.1.1b Poster displayed at the United Kingdom Radiology Congress (UKRC), Liverpool, UK 2016

8th International Conference on Children's Bone Health

10–13 June 2017, Würzburg, Germany

ICCBH
10-13 June 2017
WÜRZBURG, GERMANY

Assessment of a semi-automated software program for the identification of vertebral fractures in children

F.F. Alqahtani a,b,**, F. Messina c, E. Kruger d, H. Gill e, M. Ellis e, I. Lang d, P. Broadley d, A.C. Offiah a ,d

a Academic Unit of Child Health, University of Sheffield, Sheffield, UK
b Department of Radiological Sciences, College of Applied Medical Sciences, Najran University, Najran, Saudi Arabia
c School of Health and Related Research, University of Sheffield, Sheffield, d Radiology Department, Sheffield Children's NHS Foundation Trust, Sheffield, UK
e Sheffield Medical School, University of Sheffield, Sheffield, UK
* Article in press Clinical Radiology. (2017) doi http://dx.doi.org/10.1016/j.crad.2017.04.010
** Contact: ffmalqahtani1@Sheffield.ac.uk

Abstract: (Your abstract must use **Normal** style and must fit in this box. Your abstract should be no longer than 350 words or 280 if a table or figure is included. The box will 'expand' over 2 pages as you add text/diagrams into it.)

Purpose:
We aimed to assess observer reliability and diagnostic accuracy in children, of a semi-automated 6-point technique developed for vertebral fracture diagnosis in adults, which records percentage loss of vertebral body height.

Methods:
Reading 137 spine radiographs of children and adolescents, diagnostic accuracy (sensitivity, specificity and 95% confidence interval) calculations of five observers' for SpineAnalyzer were calculated. Comparison was made with a previously established consensus arrived at by three experienced pediatric radiologists using a simplified algorithm based qualitative scoring system (sABQ).

Results:
Of a total of 1781 vertebrae, 1187 (67%) were adequately visualized by 3 or more observers. Overall, 20 (15 %) patients had one or more VF (vertebral height loss 20 % or more). Interobserver agreement in vertebral readability for each vertebral level for five observers ranged from 0.05 to 0.47 (95% CI, -0.19, 0.76). Intraobserver agreement using the intraclass correlation coefficient (ICC) ranged from 0.25 to 0.61. Overall sensitivity and specificity were 18% (95% CI, 14 – 22) and 97% (95% CI, 97 – 98) respectively.
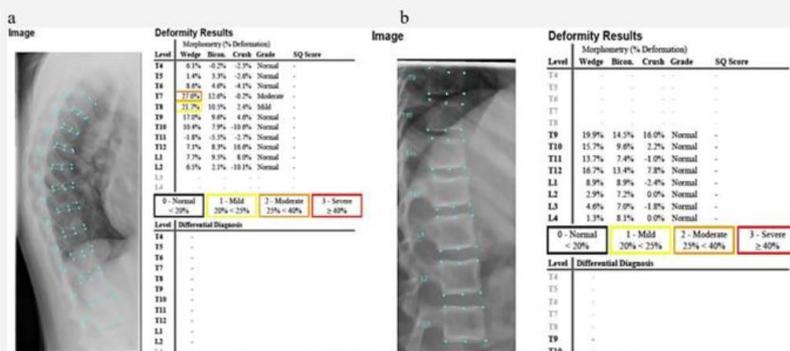
Fig. a False positive SpineAnalyzer result. Wedging of T7 and T8 as indicated by SpineAnalyzer was reported by the consensus expert panel as physiological, rather than pathological wedging
Fig. b False negative SpineAnalyzer result. T11, T12 and L2 were reported by the consensus expert panel as fractured but were scored normal by SpineAnalyzer

Conclusion:
In contrast to adults, the six-point technique assessing anterior, middle and posterior vertebral height ratios is neither satisfactorily reliable nor sensitive for VF diagnosis in children. Training of the software on pediatric images is required, in order that a pediatric standard is developed which incorporates not only specific vertebral body height ratios but also the age-related physiological changes in vertebral shape that occur throughout childhood.

Assessment of a semi-automated software program for the identification of vertebral fractures in children*

F.F. Alqahtani [a, b, **], F. Messina [c], E. Kruger [d], H. Gill [e], M. Ellis [e], I. Lang [d], P. Broadley [d], A.C. Offiah [a, d]

[a] Academic Unit of Child Health, University of Sheffield, Sheffield, UK
[b] Department of Radiological Sciences, College of Applied Medical Sciences, Najran University, Najran, Saudi Arabia
[c] School of Health and Related Research, University of Sheffield, Sheffield,
[d] Radiology Department, Sheffield Children's NHS Foundation Trust, Sheffield, UK
e Sheffield Medical School, University of Sheffield, Sheffield, UK
* Article in press Clinical Radiology. (2017) doi http://dx.doi.org/10.1016/j.crad.2017.04.010
** Contact: ffmalqahtani1@Sheffield.ac.uk

The University Of Sheffield.

1. Introduction

- The detection of one or multiple vertebral fractures (VF) - identified by a 20% reduction in vertebral body height - is indicative of bone fragility, irrespective of the reported bone mineral density (BMD)[1]
- Unlike osteoporotic fractures of the limbs, VF are typically silent and if untreated will lead to progressive loss of vertebral body height
- However if VF are diagnosed early, bisphosphonate treatment can help to treat existing and stop the occurrence of new VF
- There is significant inter and intraobserver variability in diagnosing VF in children, with a need to develop more objective methods[2]
- Semi-automatic software programmes such as SpineAnalyzer* may be the solution

*SpineAnalyzer is a semi-automated 6-point technique for VF diagnosis in adults, which records percentage loss of vertebral body height

2. Purpose

- To assess observer reliability and diagnostic accuracy of SpineAnalyzer in children

3. Materials and Methods

Image Selection

- 137 existing radiographs of children aged between 5 and 15 years, recruited into a previous study[3]

Image Interpretation

- SpineAnalyzer was used by five observers to independently analyse lateral spine radiographs as follows:
a. The reader identifies T12 (lowest visible ribs) then places a point in the centre of all vertebrae between T4 and L4; the software numbers the vertebral bodies accordingly and assigns 6 points
b. The vertebral body outline is generated by manually adjusting the 6 points (as required) to record the anterior, middle and posterior vertebral heights
c. The software generates a deformity results table (Figure 1)
- A previous consensus read by 3 paediatric radiologists using a simplified ABQ technique[3] (i.e. no software involved) served as the reference standard against which to compare the results obtained by SpineAnalyzer

Statistical Analyses

- Kappa statistics (k), intraclass correlation coefficient (ICC) and 95% confidence intervals (CI) calculated on evaluable vertebral bodies
- Sensitivity and specificity of the software tool
- R software was employed to conduct data analyses

Figure 1: Lateral thoracolumbar spine radiograph illustrating the six semi-automatically identified points used to outline the vertebral bodies and the deformity result produced by the SpineAnalyzer program

4. Results

- Mean age of the 137 subjects was 12.0 years (range, 5 to 15); 67% (92/137) were female
- Of a total of 1781 vertebrae, 1187 (67%) were adequately visualised by 3 or more observers (Figure 2)
- T4 was the most unreadable level 83/137 (61%) and the two highest visualised levels were L2 and L4; 113/137 (82%), and 111/137 (81%) respectively
- Interobserver agreement (average k score) for each vertebral level ranged from 0.054 to 0.476
- Overall sensitivity and specificity were 18% (95% CI 14 – 22) and 97% (95% CI 97 – 98) respectively
- Intra-observer agreement using ICC ranged from 0.25 to 0.61 (Table 1)
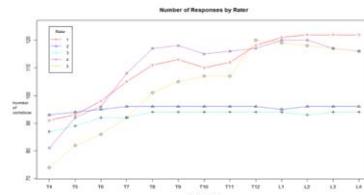
Figure 2: The number of readable vertebrae for individual observers, showing a trend towards increased readability from T to L spine

Figure 3a illustrates agreement. All 5 observers identified a severe fracture at T8

Figure 3b illustrates disagreement. T5, T6, and T7 were deemed non-evaluable by one observer, two observers called these mild, one observer moderate, while the fifth observer identified no fractures

Figure 4a False positive SpineAnalyzer result. Wedging of T7 and T8 as indicated by SpineAnalyzer was reported by the consensus expert panel as physiological, rather than pathological wedging
Figure 4b False negative SpineAnalyzer result. T11, T12 and L2 were reported by the consensus expert panel as fractured but were scored normal by SpineAnalyzer

5. Discussion

- There was relatively good readability of vertebral bodies of mid thoracic and lumbar spine but slightly limited in the upper thoracic spine. Reasons include the summation caused by intrathoracic tissues and shoulders; poor image quality; and patient positioning
- SpineAnalyzer had an overall sensitivity of only 18% and specificity of 97%
- Low sensitivity may be due to
1. The subjectivity related to the semi-automated placement of six points
2. The use of the Genant system[4] (developed from spine radiographs of osteoporotic women) to calibrate the software

6. Conclusion

- SpineAnalyzer needs training on paediatric radiographs
- Observers need training on point placement
- The six-point approach may not be reliable enough for VF diagnosis in children
- Development of specific paediatric software and normative values (incorporating age-related physiological variation in children) is required

	Sensitivity	95% CI	Specificity	95% CI	Kappa Score (k)	95% CI	ICC	SD
T4	18.20%	13.22 -23.18	96.27%	90.91 - 100.00	0.05	-0.19 - 0.30	0.30	0.43
T5	20.56%	11.25 -29.87	98.75%	96.33 - 100.00	0.36	0.13 - 0.58	0.61	0.01
T6	32.20%	16.70 - 47.70	96.51%	91.29 - 100.00	0.39	0.12 - 0.65	0.43	0.07
T7	17.25%	7.25 - 27.25	96.06%	92.51 - 99.60	0.24	-0.03 - 0.51	0.56	0.04
T8	22.80%	12.21 - 33.3	95.70%	92.56 - 98.84	0.41	0.11 - 0.72	0.54	0.02
T9	18.22%	6.39 - 30.04	96.25%	96.90 - 99.61	0.23	-0.07 - 0.53	0.38	0.05
T10	10.17%	3.40 - 16.94	98.10%	96.43 -99.77	0.17	-0.12 - 0.46	0.33	0.12
T11	19.59%	9.11 - 29.89	95.62%	92.40 - 98.84	0.23	-0.05 - 0.52	0.25	0.07
T12	24.74%	13.23 - 36.24	96.04%	92.98 - 99.10	0.47	0.18 - 0.76	0.48	0.18
L1	17.30%	9.97 - 24.67	98.32%	95.88 - 100.00	0.34	0.07 - 0.62	0.52	0.07
L2	13.70%	3.61 - 23.78	98.59%	96.57 - 100.00	0.34	-0.02 - 0.70	0.54	0.06
L3	7.93%	2.41- 13.45	99.06%	97.87 - 100.00	0.39	-0.02 - 0.80	0.46	0.01
L4	10.30%	3.41 - 17.20	99.32%	98.56 - 100.00	0.28	-0.15 - 0.67	0.43	0.06

Table 1: Sensitivity, specificity, interobserver (kappa) and intraobserver (ICC) reliability of SpineAnalyzer for vertebral fracture diagnosis in children

References:
1. Bishop, N., Arundel, P., Clark, E, Dimitri, P., Farr, J, Jones, G., Makitie, O., Munns, C. F. & Shaw, N. 2014. Fracture Prediction and the Definition of Osteoporosis in Children and Adolescents: The ISCD 2013 Pediatric Official Positions. J Clin Densitom, 17, 275-280
2. Adiotomre E, Summers L, Allison A, et al. Diagnostic accuracy of DXA compared to conventional spine radiographs for the detection of vertebral fractures in children. Eur Radiol. 2016:1-12..
3. Adiotomre, E., Summers, L., Allison, A., Walters, S. J., Digby, M., Broadley, P., Lang, I. & Offiah, A. C. 2016. Diagnosis of vertebral fractures in children: is a simplified algorithm-based qualitative technique reliable? Pediatr Radiol, 46, 680-688
4. Genant, H. K., Wu, C. Y., Vankuijk, C. & Nevitt, M. C. 1993. Vertebral fracture assessment using a semiquantitative technique. J Bone Miner Res, 8, 1137-1148

NAJRAN UNIVERSITY

NHS National Institute for Health Research

Poster 7.1.2b Poster displayed at the 8th International Conference on Children's Bone Health, Wurzburg, Germany 2017

ICCBH

8TH INTERNATIONAL CONFERENCE
ON CHILDREN'S BONE HEALTH

New Investigator Award

PRESENTED TO:

Fawaz Alqahtani

ON THE OCCASION OF THE

8th International Conference on Children's Bone Health

10-13 June 2017, Würzburg, Germany

Award 7.1.2c New Investigator Award at the 8th International Conference on Children's Bone Health, Wurzburg, Germany 2017

Abstract Submission Form

Surname: Alqahtani	First Name: Fawaz	Title (e.g. Dr, Ms, Mr):
Job Title: PhD student	**Dept & Institution:** Academic Unit of Child Health Department of Oncology and Metabolism University of Sheffield	

Please indicate if you would prefer an Oral or Poster Presentation, or either:

Poster ☐

Oral ☐

Either ☐

Email: ffmalqahtani1@sheffield.ac.uk   **Telephone:**

Abstract Title: Evaluation of a 33-point software program for the identification of vertebral fractures in children

Authors: Fawaz F Alqahtani, Fabrizio Messina, Amaka C Offiah

Abstract:

Purpose:

The aim of this study was to measure observer reliability and diagnostic accuracy in children, of a semi-automated program using a 33-point technique (Avert™) developed for vertebral fracture (VF) diagnosis in adults, which records percentage loss of vertebral body height and to compare Avert with previous results of a 6-point technique (SpineAnalyzer™)

Methods:

Lateral spine radiographs (RA) and dual-energy lateral vertebral assessment (VFA) images of 50 children and adolescents were analysed using two different programs (SpineAnalyzer™ and Avert™; 50 RA analysed with SpineAnalyzer (Group 1), 50 RA analysed with Avert (Group 2) 50 VFA analysed with Avert (Group 3). Diagnostic accuracy (sensitivity, specificity) was calculated by comparing with a previously established consensus arrived at by three experienced pediatric radiologists, using a simplified algorithm based qualitative scoring system. Levels of agreement were calculated using Cohen's kappa

Results:

Overall sensitivity of Groups 1, 2 and 3 was 26%, 37% and 35% and specificity was 98%, 94% and 92% respectively. Fair agreement was found between different modalities/different software; between Groups 2 and 3 kappa 0.29 (95% CI, 0.06 – 0.53), between Groups 1 and 3 kappa 0.35 (95% CI, 0.12 – 0.59). Moderate agreement was noted between identical modalities; Groups 1 and 2 kappa 0.55 (95% CI, 0.28 – 0.82)

Conclusion:

This study demonstrates that the 33-point technique has slightly higher accuracy for the representation of vertebral morphometry in children when compared to the 6-point technique. However, neither Avert nor SpineAnalyzer are satisfactorily reliable for VF diagnosis in children. Therefore, training of either or both software programmes on pediatric images is required

Abstract 7.1.3a Abstract submitted and accepted for the Mellanby Centre Annual Research Day, Sheffield, UK 2017

Evaluation of a 33-point software program for the identification of vertebral fractures in children

F.F. Alqahtani [a, b *], F. Messina [c], A.C. Offiah [a, d]
[a] Academic Unit of Child Health, University of Sheffield, Sheffield, UK
[b] Department of Radiological Sciences, College of Applied Medical Sciences, Najran University, Najran, Saudi Arabia
[c] School of Health and Related Research, University of Sheffield, Sheffield, UK
[d] Radiology Department, Sheffield Children's NHS Foundation Trust, Sheffield, UK
* Contact: ffmalqahtani1@sheffield.ac.uk

The University Of Sheffield.

1. Introduction

- One or multiple vertebral fractures (VF) - identified by a 20% reduction in vertebral body height – indicates bone fragility[1]
- VF are typically silent and if untreated may lead to progressive spinal deformity
- If diagnosed early, bisphosphonate treatment is beneficial
- There is significant inter and intraobserver variability in diagnosing VF in children[2]
- Semi-automatic software programmes e.g. Avert™* and SpineAnalyzer™** may be the solution

*Avert™ is a semi-automated 33-point technique for VF diagnosis in adults, which records percentage loss of vertebral body height
** SpineAnalyzer™ is a semi-automated 6-point technique for VF diagnosis in adults, which records percentage loss of vertebral body height

2. Purpose

- To measure observer reliability and diagnostic accuracy of Avert™ in children
- To compare Avert with previous results of SpineAnalyzer™ [3] in children

3. Materials and Methods

Image Selection

- 50 existing radiographs (XR) and dual energy X-ray absorptiometry (DXA) scans performed on the same day on children aged between 5 and 18 years
- All images were analysed using SpineAnalyzer™ and Avert™

Image Interpretation

- Both programs were used by a single observer (FFA) to analyse images as follows:

a. T12 identified (lowest visible ribs) then a point is placed at the centre of all vertebrae between T4 and L4; the software numbers the vertebral bodies accordingly and assigns 33 points for Avert™ and 6 points for SpineAnalyzer™ (Figure 1)

1a **1b**
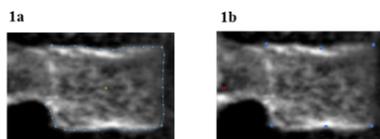
Figure 1: The 33 (1a, Avert™) and 6 (1b, SpineAnalyzer™) manually positioned points are used to outline the vertebral bodies from T4 to L4

b. The vertebral body outline is generated by manually adjusting the points (as required) to record the anterior, middle and posterior vertebral heights
c. The software generates a deformity results table
d. A previous consensus read by 3 paediatric radiologists using a simplified ABQ technique[4] (i.e. no software involved) served as the gold standard against which to compare the results obtained by Avert™ and SpineAnalyzer™

Statistical Analysis

- Sensitivity and specificity of the software tool
- Kappa statistics (k) and 95% confidence intervals (CI) calculated on evaluable vertebral bodies
- R software was employed to conduct data analysis

4. Results

- Mean (range) age of the 50 subjects was 12.0 (5 to 18) years; 17 (34%) were female
- Interobserver reliability, sensitivity and specificity of Avert™ and SpineAnalyzer™ for all vertebrae from T4 to L4 are shown in Figures 2, 3 and 4 respectively and in Table 1. Agreement between the two programmes is shown in Figure 5

KEY:

- 50 DXA analysed with Avert™ (----)
- 50 XR analysed with Avert™ (----)
- 50 DXA analysed with SpineAnalyzer™ (----)
- 50 XR analysed with SpineAnalyzer™ (----)
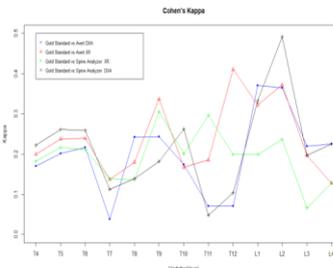
Figure 2: Interobserver (Kappa) reliability
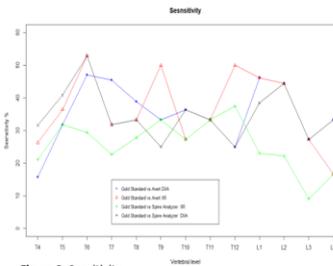
Figure 3: Sensitivity
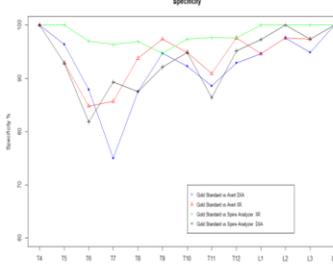
Figure 4: Specificity

Table 1. Overall sensitivity, specificity and interobserver (kappa) reliability of Avert™ and SpineAnalyzer™ for VF diagnosis in children

	Sensitivity	95% CI	Specificity	95% CI	Kappa	95% CI
Avert DXA	0.35	(0.17,0.54)	0.92	(0.77,1.00)	0.20	(0.01,0.39)
Avert XR	0.37	(0.17,0.57)	0.94	(0.81,1.00)	0.24	(0.06,0.42)
SpineAnalyzer DXA	0.35	(0.18,0.52)	0.93	(0.90,1.00)	0.22	(0.02,0.42)
SpineAnalyzer XR	0.26	(0.09,0.42)	0.98	(0.79,1.00)	0.19	(0.04,0.35)

5a **5b**
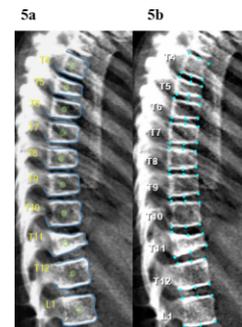
Figure 5: Lateral spine DXA analysed by Avert™ (5a) and SpineAnalyzer™ (5b) which illustrates:
- Agreement: Both programmes identified a severe fracture at T11; moderate fractures at T5 and T6; mild fractures at T7 and T8
- Disagreement: T9 identified as mild fracture by Avert™ but normal by SpineAnalyzer™
- Gold standard values: T5, T7, T8 and T9 classified as mild fracture; T6 as normal and T11 as moderate fracture

5. Discussion

- Avert™ has slightly higher accuracy for diagnosis of VF in children compared to SpineAnalyzer™
- Low sensitivity of both programmes may be due to
1. The subjectivity related to the semi-automated placement of points
2. The use of the Genant system[5] (developed from spine radiographs of osteoporotic women) for calibration

6. Conclusion

- Neither Avert™ nor SpineAnalyzer™ is satisfactorily reliable for VF diagnosis in children
- Development of specific paediatric software and normative values (incorporating age-related physiological variation in children) is required

7. Acknowledgments

National Institute for Health Research — NAJRAN UNIVERSITY

References:
1. Bishop, N., Arundel, P., Clark, E, Dimitri, P., Farr, J, Jones, G., Makitie, O., Munns, C. F. & Shaw, N. 2014. Fracture Prediction and the Definition of Osteoporosis in Children and Adolescents: The ISCD 2013 Pediatric Official Positions. J Clin Densitom, 17, 275-280
2. Adiotomre E, Summers L, Allison A, et al. 2016. Diagnostic accuracy of DXA compared to conventional spine radiographs for the detection of vertebral fractures in children. Eur Radiol, 27, 2188-2199
3. F.F. Alqahtani, F. Messina, E. Kruger, H. Gill, M. Ellis, I. Lang, P. Broadley, A.C. Offiah. 2017. Evaluation of a semi-automated software program for the identification of vertebral fractures in children. Clin Radiol, 72, 904-920
4. Adiotomre, E., Summers, L., Allison, A., Walters, S. J., Digby, M., Broadley, P., Lang, I. & Offiah, A. C. 2016. Diagnosis of vertebral fractures in children: is a simplified algorithm-based qualitative technique reliable? Pediatr Radiol, 46, 680-688
5. Genant, H. K., Wu, C. Y., Vankuijk, C. & Nevitt, M. C. 1993. Vertebral fracture assessment using a semiquantitative technique. J Bone Miner Res, 8, 1137-1148

Poster 7.1.3b Poster displayed at the Mellanby Centre Annual Research Day, Sheffield, UK 2017

Presentation Details
Control #: 4559
Presentation Time: talk: 6 minutes + 2 minutes discussion
Presentation Title: Assessment of a 33-point software programme for the identification of vertebral fractures in children
Presentation Number: B-0158
Author Block: F. Alqahtani, F. Messina, A.C. Offiah; *Sheffield/GB*
Disclosure Block: F. Alqahtani: None. **F. Messina:** None. **A.C. Offiah:** None.
Topic: Paediatric
Session Number: SS 212
Session Title: Neuro imaging in paediatrics
Session Date/Time: 2/28/2018 10:30:00 AM - 2/28/2018 12:00:00 PM
Room: 18/M 2

Please **check the presentation details and your abstract** and inform us (scientific.programme@ecr.org) of **any necessary corrections by January 8, 2018** at the latest!

Below is a view of your accepted abstract:
Purpose: To measure observer reliability and diagnostic accuracy in children, of a semi-automated programme using a 33-point technique (Avert™) developed for vertebral fracture (VF) diagnosis in adults, which records percentage loss of vertebral body height and to compare Avert with previous results of a 6-point technique (SpineAnalyzer™).
Methods and Materials: Lateral spine radiographs (RA) and dual-energy lateral vertebral assessment (VFA) images of 50 children were analysed using two different programmes (SpineAnalyzer™ and Avert™; 50 RA analysed with SpineAnalyzer (Group 1), 50 RA analysed with Avert (Group 2) 50 VFA analysed with Avert (Group 3). Diagnostic accuracy (sensitivity, specificity) was calculated by comparing with a previously established consensus arrived at by three experienced paediatric radiologists, using a simplified algorithm based qualitative scoring system. Levels of agreement were calculated using Cohen's kappa.
Results: Overall sensitivity of Groups 1, 2 and 3 was 26%, 37% and 35% and specificity was 98%, 94% and 92% respectively. Fair agreement was found between different modalities/different software; between Groups 2 and 3 kappa 0.29 (95% CI, 0.06 - 0.53), between Groups 1 and 3 kappa 0.35 (95% CI, 0.12 - 0.59). Moderate agreement was noted between identical modalities; Groups 1 and 2 kappa 0.55 (95% CI, 0.28 - 0.82).
Conclusion: The 33-point technique has slightly higher accuracy for the representation of vertebral morphometry in children when compared to the 6-point technique. However, neither Avert nor SpineAnalyzer are satisfactorily reliable for VF diagnosis in children. Therefore, training of either or both software programmes on paediatric images is required.

Click here to read through the guidelines for oral presentation.

You have completed the ECR 2018 confirmation process. Thank you.

Your Response: I_confirm_to_present

Questions?	**Technical Support Helpdesk**
European Society of Radiology	Phone: +1-217-398-1792
Email: manuela.gewessler@myesr.org	Email: OASIS Helpdesk

Abstract 7.1.4 Abstract submitted and accepted for the 30th European Congress of Radiology (ECR), Vienna, Austria 2018. Insights into Imaging Abstracts. DOI.org/10.1007/s13244-018-0603-8 (oral presentation)

Abstract Submission Form

Surname: Alqahtani	First Name: Fawaz	Title (e.g. Dr, Ms, Mr):
Job Title: PhD student- Radiographer	Dept & Institution: Academic Unit of Child Health Department of Oncology & Metabolism University of Sheffield	

Please indicate if you would prefer an Oral or Poster Presentation, or either:

Poster ☐

Oral ☐

Either X

Email: ffmalqahtani1@sheffield.ac.uk   Telephone:

Abstract Title: Diagnostic performance of morphometric vertebral fracture analysis (MXA) using a 33-point software programme on DXA images of children

Authors: Fawaz F Alqahtani, Nicola J Crabtree and Amaka C Offiah

Abstract:
No more than 300 words
Purpose:

The aim of this study was to evaluate the diagnostic accuracy of morphometric vertebral fracture analysis (MXA) using a 33-point software programme designed for adults, on iDXA images of children.

Methods:

Lateral spine iDXA of 420 children aged between 5 to 18 years were retrospectively recruited. The majority of subjects were considered high risk for vertebral fracture. Vertebral fracture assessment of an expert paediatric radiologist using the Genant's semiquantitative scoring system served as the gold standard. All 420 DXA scans were analysed using semi-automated software (33-point morphometry) by a radiographer. MXA of a random sample of 30 DXA were analysed by three paediatric radiologists and an experienced paediatric clinical scientist. Diagnostic accuracy (sensitivity, specificity, false positive, false negative and 95% confidence interval) was calculated. Inter and intra-observer agreement levels were calculated using kappa statistics.

Results:

Overall sensitivity, specificity, false positive and false negative rates for the radiographer analysing 420 DXA were 80% 90%, 10% and 20% respectively. Overall sensitivity, specificity, false positive and false negative rates for the 4 additional observers were 89%, 79%, 21% and 11% respectively. Moderate to good agreement was found between MXA and the gold standard [kappa ranged from 0.41 to 0.66 (95% CI: 0.29 - 0.82)]. By contrast, only moderate inter and intra-observer agreement was noted between all raters for MXA, Fleiss' kappa (κ) was 0.44 (95% CI: 0.48 - 0.99) and 0.60 (95% CI: 0.48 - 0.75) respectively.

Conclusion:

MXA using a 33-point technique developed for adults is not a reliable method for the identification of mild vertebral fractures in children. New methods to facilitate the detection of vertebral fractures in children are needed. In particular, a paediatric standard is required which incorporates not only specific vertebral body height ratios but also the age-related physiological changes in vertebral shape that occur throughout childhood.

Abstract7.1.5a Abstract submitted and accepted for the Mellanby Centre Annual Research Day, Sheffield, UK 2018

Diagnostic performance of morphometric vertebral fracture analysis (MXA) in children using a 33-point software programme

F.F. Alqahtani [a, b*], N. J. Crabtree [c], P. Bromiley [d], T. Cootes [d], I. Lang [e], P. Broadley [e], A.C. Offiah [a, e]

[a] Academic Unit of Child Health, University of Sheffield, Sheffield, UK
[b] Department of Radiological Sciences, College of Applied Medical Sciences, Najran University, Najran, Saudi Arabia
[c] Department of Endocrinology and Diabetes, Birmingham Children's Hospital, Birmingham, UK
[d] Imaging Sciences Research Group, University of Manchester, Manchester, UK
[e] Radiology Department, Sheffield Children's NHS Foundation Trust, Sheffield, UK
* Contact: ffmalqahtani1@sheffield.ac.uk

The University Of Sheffield.

1. Introduction

- The detection of one or multiple vertebral fractures (VF) - identified by a 20% reduction in vertebral body height - is indicative of bone fragility, irrespective of the reported bone mineral density (BMD)[1]
- Unlike osteoporotic fractures of the limbs, VF are typically silent and if untreated may lead to progressive loss of vertebral body height
- However if VF are diagnosed early, bisphosphonate treatment can help to treat existing and stop the occurrence of new VF
- There is significant inter and intraobserver variability in diagnosing VF in children, with a need to develop more objective methods[2]
- Semi-automatic software programmes such as Avert* may be the solution

  *Avert is a semi-automated 33-point technique for VF diagnosis in adults, which records percentage loss of vertebral body height

2. Purpose

- To assess the diagnostic accuracy and observer reliability of morphometric VF analysis (MXA) using a 33-point software programme designed for adults (Avert), on iDXA images of children.

3. Materials and Methods

Image Selection

- 420 existing iDXA lateral spine images of children aged between 5 and 18 years were selected
- Images were analysed independently by five observers (three paediatric radiologists R1, R4, R5, a radiographer R2, and clinical scientist R3), with each observer being blinded to the other evaluations
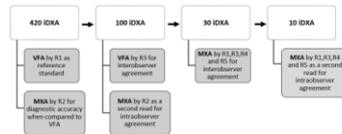
Figure 1. Flow chart summarising the reporting pathway

Image Interpretation

VFA (R1,R3)

- Identification of VF was performed on the DXA by visually assessing the T4 to L4 vertebrae, relying on semi-quantitaive (SQ) method based on Genant's classification (i.e., with no software involved)[3]

MXA (R1-R5)

- Avert was used by five observers to independently analyse lateral iDXA scans as follows:
a) The reader identifies T12 (a marker at T12 for all images was placed for consistency between observers), then places a point in the centre of all vertebrae between T4 and L4; the software numbers the vertebral bodies accordingly and assigns 33 points (Figure 2)
b) The vertebral body outline is generated by manually adjusting the 33 points (as required) to record the anterior, middle and posterior vertebral heights
c) The software generates deformity results

Statistical Analyses

- Kappa statistics (k) and 95% confidence intervals (CI) calculated on evaluable vertebral bodies
- Sensitivity and specificity of the software tool
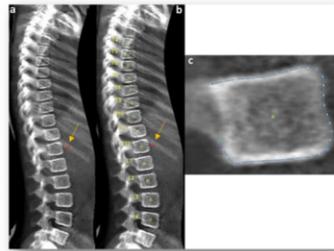
Figure 2 Technique used to perform semi-automated quantitative morphometry measurements (Avert). **a)** Lateral iDXA scan of the entire spine of a 9-year-old female with osteogenesis imperfecta; **b)** identified vertebral bodies from T4 to L4; **c)** 33 points placed to outline T12. The arrow points to the T12 level

4. Results

- We included 420 lateral iDXA scans in children aged between 5 and 18 years; 210 (50 %) were male; 380 (90%) had osteogenesis imperfecta
- Among the 420 subjects, 191 (45%) had no fracture by the gold standard visual SQ method, while mild, moderate and severe fractures were identified in 98 (23%), 67 (16%), and 29 (7%) subjects respectively
- Isolated physiological wedging (with no fracture) was identified in 35 (8%) children

- _Diagnostic accuracy of MXA (420 iDXA)_

Figures 3a and 3b show the number, severity, and shape of VFs by the two methods at the vertebral and subject levels respectively, as well as the physiological wedges identified by VFA.

3a At vertebral level
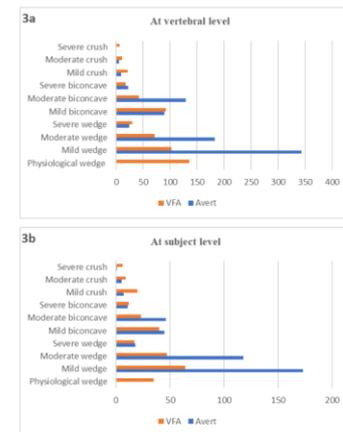
3b At subject level

Figure 3. Number of vertebral fracture shapes identified using both techniques (a) at vertebral level and (b) at subject level (note that Avert does not have the ability to diagnose physiologically wedged vertebrae)

Table 1. Diagnostic accuracy of MXA for detecting vertebral fractures (n = 420 patients, 5460 vertebra)

	Any fracture	Clinical fracture*	Mild fracture**
Sensitivity	80%	87%	46%
Specificity	90%	95%	92%
Kappa (95% CI)	0.37 (0.30 – 0.43)	0.43 (0.35 – 0.51)	0.41 (0.27 – 0.5

*Clinical fracture = moderate and severe (≥26%) vertebral height reduction;
** Mild fracture = 21% to 25% vertebral height reduction

- _Intraobserver agreement of MXA (100 iDXA)_

There was fair to excellent intraobserver agreement, with kappa ranging from 0·49 to 0·87 (95% CI 0·37 – 0·98)

- _Observer agreement of MXA (30 iDXA)_

The overall average sensitivity, specificity, false positive, and false negative rates for the four observers were 89%, 79%, 21%, and 11%, but when only mild fractures were considered, the averages were 36%, 86%, 14%, and 64% respectively.

Table 2. Summary of inter and intraobserver agreement for MXA (n=30)

	Agreement across four observers (Kappa)			
	Interobserver agreement	(95% CI)	Intraobserver agreement	(95% CI)
Any fracture	0.44	(0.31 – 0.56)	0.55 – 0.68	(0.45 – 0.7
Clinical fracture	0.52	(0.40 – 0.61)	0.59 – 0.89	(0.50 – 0.9
Mild fracture	0.21	(0.09 – 0.33)	0.51 – 0.67	(0.41 – 0.7

- _Observer agreement of VFA (100 iDXA)_

Interobserver agreement between two expert observers was fair to good for VFA, with kappa ranging from 0·29 to 0·76 (95% CI: 0·17 – 0·88)

5. Discussion

- Results demonstrate that MXA is only as good as VFA in identifying severe VFs with reduced diagnostic accuracy for detecting mild VFs
- Our results suggest that a large contributory factor is the inability of the software to distinguish normal physiological wedging (i.e. developmental morphological variability that occurs throughout childhood) from mild fractures, particularly in the thoracic region
- Another major limitation of MXA is the inability of the software programme to identify fractures when height loss is below 20%, as identified in 32 subjects (8%)
- This inability to differentiate normal physiological wedging from fracture also accounts for low diagnostic accuracy of VFA

6. Conclusion

- MXA reaches only moderate agreement when compared to the visual SQ VFA technique, with fair to moderate inter and intraobserver agreement
- In order to facilitate the detection of mild VF in children, a paediatric standard is required which not only incorporates specific vertebral body height ratios but also the age-related physiological changes in vertebral shape that occur throughout childhood

References:
1. Bishop, N., Arundel, P., Clark, E, Dimitri, P., Farr, J, Jones, G., Makitie, O., Munns, C. F. & Shaw, N. 2014. Fracture Prediction and the Definition of Osteoporosis in Children and Adolescents: The ISCD 2013 Pediatric Official Positions. J Clin Densitom, 17, 275-280
2. Alqahtani, F.F. & Offiah, A.C. Diagnosis of osteoporotic vertebral fractures in children. Pediatr Radiol (2018). https://doi.org/10.1007/s00247-018-4279-5
3. Genant, H. K., Wu, C. Y., Vankuijk, C. & Nevitt, M. C. 1993. Vertebral fracture assessment using a semiquantitative technique. J Bone Miner Res, 8, 1137-1148

Poster 7.1.5b Poster displayed at the Mellanby Centre Annual Research Day, Sheffield, UK 2018

7.2 Appendix 2 – Publications

Evaluation of a semi-automated software program for the identification of vertebral fractures in children

F.F. Alqahtani [a,b,*], F. Messina [c], E. Kruger [d], H. Gill [e], M. Ellis [e], I. Lang [d], P. Broadley [d], A.C. Offiah [a,d]

[a] Academic Unit of Child Health, University of Sheffield, Sheffield, UK
[b] Department of Radiological Sciences, College of Applied Medical Sciences, Najran University, Najran, Saudi Arabia
[c] School of Health and Related Research, University of Sheffield, Sheffield, UK
[d] Radiology Department, Sheffield Children's NHS Foundation Trust, Sheffield, UK
[e] Sheffield Medical School, University of Sheffield, Sheffield, UK

AIM: To assess observer reliability and diagnostic accuracy in children, of a semi-automated six-point technique developed for vertebral fracture (VF) diagnosis in adults, which records percentage loss of vertebral body height.

MATERIALS AND METHODS: Using a semi-automated software program, five observers independently assessed T4 to L4 from the lateral spine radiographs of 137 children and adolescents for VF. A previous consensus read by three paediatric radiologists using a simplified algorithm-based qualitative technique (i.e., no software involved) served as the reference standard.

RESULTS: Of a total of 1,781 vertebrae, 1,187 (67%) were adequately visualised according to three or more observers. Interobserver agreement in vertebral readability for each vertebral level for five observers ranged from 0.05 to 0.47 (95% CI: -0.19, 0.76). Intra-observer agreement using the intraclass correlation coefficient (ICC) ranged from 0.25 to 0.61. The overall sensitivity and specificity were 18% (95% CI: 14–22) and 97% (95% CI: 97–98), respectively.

CONCLUSION: In contrast to adults, the six-point technique assessing anterior, middle, and posterior vertebral height ratios is neither satisfactorily reliable nor sensitive for VF diagnosis in children. Training of the software on paediatric images is required in order to develop a paediatric standard that incorporates not only specific vertebral body height ratios but also the age-related physiological changes in vertebral shape that occur throughout childhood.

Introduction

* Guarantor and correspondent: F. F, Alqahtani, Room C7, C Floor, Academic Unit of Child Health, Department of Oncology and Metabolism, University of Sheffield, Medical School, Stephenson Wing, Sheffield Children's Hospital, Western Bank, Sheffield S10 2TH, UK. Tel.: 0114 271 7228.
E-mail address: ffmalqahtani1@sheffield.ac.uk (F.F. Alqahtani).

Fractures are common in childhood and repeated fractures reflect the interacting effects of low bone mineral density (BMD) and/or physical activity.[1] Vertebral fractures (VFs) are a relatively common type of osteoporotic fracture.

Article 7.2.1 Title page of original article published in Clinical Radiology, 2017. DOI.org/10.1016/j.crad.2017.04.010

REVIEW

CrossMark

Diagnosis of osteoporotic vertebral fractures in children

Fawaz F. Alqahtani[1,2] · Amaka C. Offiah[1,3]

Abstract

Osteoporosis is a generalised disorder of the skeleton with reduced bone density and abnormal bone architecture. It increases bone fragility and renders the individual susceptible to fractures. Fractures of the vertebrae are common osteoporotic fractures. Vertebral fractures may result in scoliosis or kyphosis and, because they may be clinically silent, it is imperative that vertebral fractures are diagnosed in children accurately and at an early stage, so the necessary medical care can be implemented. Traditionally, diagnosis of osteoporotic vertebral fractures has been from lateral spine radiographs; however, a small number of studies have shown that dual energy x-ray absorptiometry is comparable to radiographs for identifying vertebral fractures in children, while allowing reduced radiation exposure. The diagnosis of vertebral fractures from dual energy x-ray absorptiometry is termed vertebral fracture assessment. Existing scoring systems for vertebral fracture assessment in adults have been assessed for use in children, but there is no standardisation and observer reliability is variable. This literature review suggests the need for a semiautomated tool that (compared to the subjective and semiquantitative methods available) will allow more reliable and precise detection of vertebral fractures in children.

Keywords Children · Dual energy x-ray absorptiometry · Diagnostic scoring system · Osteoporosis · Vertebral fracture · Vertebral fracture assessment

Introduction

Fractures are common in childhood. About one-third of children in the United Kingdom will have at least one fracture during their childhood [1]. Osteoporotic vertebral fractures are increasingly recognised in children with either primary (e.g., osteogenesis imperfecta) [2] or secondary low bone mineral density (e.g., acute lymphoblastic leukaemia, inflammatory bowel disease and glucocorticoid use) [3, 4]. Nearly 1 in 5 children with a rheumatological condition will have a vertebral fracture [5] and rates are similar or even higher in other conditions, e.g., 16% in acute lymphoblastic leukaemia [6], up to 75% in Duchenne muscular dystrophy [7] and up to 100% in severe forms of osteogenesis imperfecta (personal experience of the senior author). Outside the context of major trauma, vertebral fractures in children indicate pathological bone fragility and precise and early diagnosis is imperative so appropriate medical care can be initiated.

Techniques to detect and analyse vertebral fractures in clinical and/or research practice include conventional radiography, computed tomography (CT), magnetic resonance imaging (MRI) and dual energy x-ray absorptiometry. Traditionally, the most common method for diagnosing vertebral fractures is x-ray, although dual energy x-ray absorptiometry has now been shown to diagnose vertebral fractures with the advantage of also determining bone mineral density [8]. Vertebral fracture assessment is the term given to the diagnosis of vertebral fractures from dual energy x-ray absorptiometry scans [9]. This technology is more or less in routine clinical use in adults, complemented by validated scoring systems [10, 11]. Conversely, vertebral fracture assessment is less widely

✉ Fawaz F. Alqahtani
   ffmalqahtani1@sheffield.ac.uk

[1] Academic Unit of Child Health,
    Department of Oncology and Metabolism,
    Medical School, Damer Street Building,
    Sheffield Children's Hospital,
    University of Sheffield,
    Western Bank, Sheffield S10 2TH, UK

[2] Department of Radiological Sciences,
    College of Applied Medical Sciences,
    Najran University,
    Najran, Kingdom of Saudi Arabia

[3] Radiology Department,
    Sheffield Children's NHS Foundation Trust,
    Sheffield, UK

🖄 Springer

Article 7.2.2 Title page of review article published in Pediatric Radiology, 2018.
DOI.org/10.1007/s00247-018-4279-5

7.3 Appendix 3 – Permissions

Permission 7.3.1 Permission to reuse the review paper (Chapter Two) in the thesis

Permission 7.3.2 Permission to reuse the above paper (Chapter Three) in the thesis

REQUEST TO SUBMIT AN ALTERNATIVE FORMAT THESIS

Students should use this form to request permission to submit a doctoral thesis which contains sections which are in a format suitable for submission for publication in a peer-reviewed journal (or other appropriate outlet for academic research), alongside more traditional thesis chapters. This avoids the need for students to spend time rewriting publications into thesis chapters, as well as enhancing their writing for publication skills.

Requests to submit an alternative format thesis must be supported by the student's supervisory team and also require approval from the relevant Faculty and must therefore be made well in advance of the student submitting their thesis and no later than the end of the student's 3rd year of study.

Students wishing to submit in this format must read the guidance on structure of an alternative format thesis at: https://www.sheffield.ac.uk/rs/code/altformat

SECTION 1: STUDENT'S DETAILS	
NAME: Fawaz Alqahtani	REGISTRATION NUMBER: 140261343
DEPARTMENT: Oncology and Metabolism	FACULTY: Medicine, Dentistry & Health

PROPOSED THESIS TITLE:
Semi-Automated Vertebral Fracture Assessment In Children

REASON FOR THE REQUEST: Please use this space to provide details of the published papers that will be incorporated into your thesis and how your thesis will be structured (continue on a separate sheet, if required)

I would like to submit my PhD thesis as an alternative format for the following reasons:

1- To improve my academic writing skills for publication during my PhD time
2- Helping me to avoid duplication and save time by not having to convert publications into thesis chapters

I am planning to submit my thesis as publication papers as follows:
Chapter 1:
General Introduction and Background
Chapter 2:
Vertebral Fracture Diagnosis in Children: A Literature Review. (Aiming to submit to Pediatric Radiology by March 31, 2018)
Chapter 3:
Evaluation of a semi-automated software program for the identification of vertebral fractures in children. (published 13 May 2017, Clinical Radiology https://doi.org/10.1016/j.crad.2017.04.010
Chapter 4:
Comparison between two software programs using two modalities (Radiographs & DXA) for the identification of vertebral fractures in children. (Accepted for oral presentation at an international conference in March 2018, aiming to submit to Clinical Radiology by July 31, 2018)
Chapter 5:
Training the existing adult software tool (AVERT) on existing paediatric spine DXA (Aiming to submit for publication by October 31, 2018).
Chapter 6:
Overall Discussion and Conclusion

SIGNATURE: *Fawaz Algahtani*	DATE 30th January 2018

SECTION 3: FACULTY CONSIDERATION		
NAME:	FACULTY:	
COMMENTS ON DECISION:		
☑	I approve this request for submission of an alternative format thesis	☐	I do not approve this request for submission of an alternative format thesis
SIGNATURE: *Martina Daly.*	DATE: *7/2/2018·*	

Permission 7.3.3 Permission to submit the thesis as alternative format