Observer agreement in pediatric semi-quantitative vertebral fracture diagnosis

Kerry Siminoski,

Department of Radiology and Diagnostic Imaging and Division of Endocrinology and Metabolism,
Department of Medicine, University of Alberta, 6628-123 Street, Edmonton, Alberta, Canada T6H
3T6

Brian Lentle,

Department of Radiology, BC Children's Hospital, University of British Columbia 4500 Oak St,
Vancouver, British Columbia, Canada V6H 3N1

Mary-Ann Matzinger,

Department of Diagnostic Imaging, Children's Hospital of Eastern Ontario, University of Ottawa
401 Smyth Road, Ottawa, Ontario, Canada K1H 8L1

Nazih Shenouda,

Department of Diagnostic Imaging, Children's Hospital of Eastern Ontario, University of Ottawa
401 Smyth Road, Ottawa, Ontario, Canada K1H 8L1

Leanne M. Ward, and

Department of Pediatrics, Children's Hospital of Eastern Ontario, University of Ottawa 401 Smyth
Road, Ottawa, Ontario, Canada K1H 8L1

the Canadian STOPP Consortium

Abstract

Background—The Genant semi-quantitative (GSQ) method has been a standard procedure for diagnosis of vertebral fractures in adults, but has only recently been shown to be of clinical utility in pediatrics. Observer agreement using the GSQ method in this age group has not been described.

Objective—To evaluate observer agreement on vertebral readability and vertebral fracture diagnosis using the GSQ method in pediatric vertebral morphometry.

Materials and methods—Spine radiographs of 186 children with acute lymphoblastic leukemia were evaluated independently by three radiologists using the same GSQ methodology as in adults. A subset of 100 radiographs was evaluated on two occasions.

Results—An average of 4.7% of vertebrae were unreadable for the three radiologists. Intraobserver Cohen's kappa (κ) on readability ranged from 0.434 to 0.648 at the vertebral level and from 0.416 to 0.611 at the patient level, while interobserver κ for readability had a range of 0.330 to 0.504 at the vertebral level and 0.295 to 0.467 at the patient level. Intraobserver κ for the

presence of vertebral fracture had a range of 0.529 to 0.726 at the vertebral level and was 0.528 to 0.767 at the patient level. Interobserver κ for fracture at the vertebral level ranged from 0.455 to 0.548 and from 0.433 to 0.486 at the patient level.

Conclusion—Most κ values for both intra- and interobserver agreement in applying the GSQ method to pediatric spine radiographs were in the moderate to substantial range, comparable to the performance of the technique in adult studies. The GSQ method should be considered for use in pediatric research and clinical practice.

Keywords

acute lymphoblastic leukemia; observer agreement; osteoporosis; pediatric radiology; vertebral fracture; vertebral morphometry

Introduction

There has been extensive investigation of vertebral fractures in adults over the past 20 years [1–9]. In contrast, the study of pediatric vertebral fracture is in its infancy, in part because these fractures are much less common in the general pediatric population than in adults, in part because of the lack of an accepted technique for diagnosis [10, 11]. An important impetus to adult vertebral fracture research was the development of vertebral morphometry techniques that were subsequently shown to be clinically meaningful [12]. The predominant approach used in the field of adult osteoporosis is the modified Genant semi-quantitative method (GSQ method) [1–9]. This system has been used in many of the major drug trials and in clinical studies and has been recommended for use in routine clinical practice [13–15]. The procedure involves visually assessing anterior, middle, and posterior height ratios of each vertebral body from T4 to L4 on lateral spine images, with height reductions graded on a four-point scale [1, 3]. The method is considered semi-quantitative by virtue of assigning approximate ratios within a range defined by each grade rather than using exact measurements. The advantage of this method over strict quantitative morphometry is that it utilizes radiologist expertise in assessing vertebral morphology. This is particularly important if the technique is to be applied to pediatric spine imaging, where incomplete calcification of the centrum must be considered in the evaluation [1, 16, 17].

Vertebral fractures are increasingly recognized as an important clinical consequence of osteoporosis in children and adolescents, particularly those treated with glucocorticoids [17–19]. There is not yet an accepted morphometric approach to radiological assessment, although several research groups have developed methodologies for their own use [20–22]. We have recently applied the GSQ method in children with rheumatic conditions, where GSQ method-defined vertebral fractures correlated with the presence of back pain, and in children with acute lymphoblastic leukemia (ALL), where vertebral fractures correlated with bone mineral density (BMD), metacarpal cortical area, and back pain [17, 18]. These findings suggest that the GSQ method may be an appropriate system for use in pediatric osteoporosis studies and pediatric clinical practice.

An important requirement of any diagnostic method is acceptable concordance within and between readers [23]. The reproducibility of the GSQ method in adult radiographic vertebral

morphometry has been studied in depth, with results summarized in Table 1. Intra- and interobserver agreements range from fair (Cohen's kappa statistic (κ) from 0.21 to 0.40) to almost perfect ($\kappa > 0.80$) at both the individual vertebral level (determining whether individual vertebrae are fractured) and at the patient level (determining whether a patient has one or more vertebral fractures) [1–9, 23]. As the reproducibility of the GSQ method in the pediatric population has not been described, we have evaluated observer agreement on vertebral readability and vertebral fracture diagnosis in a cohort of children with ALL, employing the same GSQ methodology and grading used in adults.

Materials and methods

Study participants

This study was approved by the Research Ethics Boards in the participating institutions and local consent processes were followed. Study subjects were children and adolescents with newly diagnosed ALL at 10 tertiary care children's hospitals enrolled in the Steroid-associated Osteoporosis in the Pediatric Population (STOPP) study. Demographic data was collected. Height, weight, and bone mineral density (BMD) values were transformed to age- and sex-matched Z-scores, which represent the number of standard deviations from the mean for a reference population of that age and sex. Height and weight Z-scores were determined using the U.S. Centers for Disease Control National Center for Health Statistics normative database and lumbar spine BMD results were determined using the Hologic12.4 normative database. Spine images were evaluated independently by three experienced radiologists (mean time since radiology accreditation 33 years, range from 20 years to 50 years). The entire cohort (N = 186) was evaluated to determine interobserver agreement, with this group referred to as the Interobserver Cohort. As assessment may be more difficult at younger ages when bone development is less advanced, the cohort was divided into two equal-sized groups based on age, the Younger Interobserver Cohort (n = 93) and the Older Interobserver Cohort (n = 93). To determine intraobserver agreement (Intraobserver Cohort), radiographs of 100 subjects were randomly chosen to be read a second time by each radiologist (with a minimum interval of at least 4 weeks and with no knowledge of the results of the first read). The Intraobserver Cohort was also divided into two equal-sized subsets (n = 50) based on age (Younger Intraobserver Cohort and Older Intraobserver Cohort).

Vertebral morphometry

Subjects underwent lateral spine radiographs using standardized technique at a median of 18 days (interquartile range, 7 to 25 days) from first chemotherapy following the diagnosis of ALL. The spine was evaluated from T4 to L4. Individual vertebrae were graded according to the GSQ method by assessing the anterior:posterior height ratio and the middle:posterior height ratio within a vertebra, and the posterior:posterior height ratio when comparing to an adjacent vertebra [1]. The methodology is illustrated in Figure 1. Examples of fractures in this patient group graded using the GSQ method have previously been published [17]. For the main analyses, fracture grades were defined as grade 1 (>20% to 25% reduction in a height ratio), grade 2 (>25 to 40% reduction), or grade 3 (>40% reduction). For some analyses, grade 0.5 was also considered, defined as reduction in a height ratio of >15% to

20%. These grades and grade definitions are the same as are used in adult studies applying the GSQ method. A record was made of vertebrae that were considered unreadable by each radiologist. Fracture morphology was not assessed in this analysis.

Statistical methods

Population characteristics are expressed as mean (SD) or median (range) for continuous variables, while discrete variables are expressed as the value (percentage frequency). Agreement within or between readers is expressed using percent simple agreement and κ with 95% confidence interval (95% CI) [23]. Weighted κ (weighted for fracture grade) was also determined in some analyses [24]. Observer agreement was assessed in two ways: at the individual vertebral level and at the patient level. For global assessment of agreement at the vertebral level for both readability and vertebral fracture, all vertebrae were combined into a single pool, giving a maximum of 2,418 comparisons. Agreement was also determined at each vertebral level for the presence of vertebral fracture, a maximum of 186 comparisons per vertebra. For determination of agreement on readability at the patient level, individuals were classified as having all vertebrae from T4 to L4 readable or to have one or more unreadable. Similarly, for vertebral fracture assessment at the patient level, patients were classified as having no fracture or having one or more fractures. The scheme of Landis & Koch was used to relate strength of agreement with ranges of κ values: no agreement, 0; slight, 0.01–0.20; fair, 0.21–0.40; moderate, 0.41–0.60; substantial, 0.61–0.80; almost perfect, 0.81–1.00 [23]. Comparison of cohort characteristics was done using Student's t-test, the Mann-Whitney-Wilcoxon test, chi-square test, or Fisher's exact test. Differences were considered significant for $p < 0.05$. Analyses were conducted using SPSS 19.0.0.1 or SAS 9.2.

Results

Study population

The entire study population was used for determination of interobserver agreement (Interobserver Cohort) and is described in Table 2. This patient group has been previously described in detail [17]. The Younger (mean age 3.4 years; range 1.3 to 5.2 years) and Older (mean age 9.6 years; range 5.3 to 17.0 years) Interobserver Cohorts differed only in age ($p < 0.001$) and weight Z-score ($p = 0.001$). The subset of subjects used for assessment of intraobserver agreement was similarly divided into Younger (mean age 5.4 years; range 1.3 to 5.3 years) and Older (mean age 10.2 years; range 5.4 to 16.6 years) Intraobserver Cohorts, which again differed in age ($p < 0.001$) and weight Z-score ($p = 0.004$).

Readability

The proportion of subjects for which all 13 vertebrae were readable averaged 75.4% (range 59.7% to 87.6% for the three radiologists). Of the 2,418 individual vertebral assessments done by each radiologist in the Interobserver Cohort, a mean of 4.7% (2.3% to 8.2%) were unreadable, ranging from 2.3% to 8.2% for the individual readers. In the Younger group a mean of 7.4% (4.1% to 12.2%) of total vertebrae were unreadable compared to 1.6% (0.8% to 2.8%) in the Older group. The upper thoracic spine had the highest unreadability (Fig. 2). From T4 to T8, the mean percentage of vertebrae that were unreadable was 10.1% (range

5.3% to 17.3%) compared to 1.2% (range 0.4% to 2.4%) from T9 to L4. K values for intraobserver and interobserver agreement on readability are shown in table 3.

Vertebral Fracture Diagnosis

Intraobserver Agreement on Vertebral Fracture Diagnosis—The intraobserver κ for the presence of vertebral fracture at the vertebral level ranged from 0.529 to 0.726 while the weighted κ ranged from 0.632 to 0.731 (Table 4). At the patient level, κ had a range of 0.528 to 0.767 and the weighted κ ranged from 0.678 to 0.778. Intraobserver agreement was similar across vertebral levels with the exception of L4, where agreement dropped off (Fig. 3A). When reproducibility of fracture diagnosis was examined by age group, there was little difference between the two groups (Table 5). To determine whether intraobserver agreement was dependent on fracture grade, different outcome dichotomies were assessed, with fracture definition thresholds ranging from grade 0.5 to grade 3 (Figure 4A, B). There was little change in agreement across fracture grades at either the vertebral or patient levels

Interobserver Agreement on Vertebral Fracture Diagnosis—The interobserver κ value for the presence of fracture at the vertebral level ranged from 0.455 to 0.548 and the weighted κ from 0.481 to 0.568 (Table 4). At the patient level the κ ranged from 0.433 to 0.486 while the weighted κ ranged from 0.525 to 0.592. Agreement was better in the lower spine than the upper spine (Fig. 3B). For T4 to T8 the κ ranged from 0.289 to 0.517 compared to 0.531 to 0.692 for T9 to L4. When reproducibility of fracture diagnosis was examined by age group, κ values were similar and overlapping for the two groups for each radiologist pair (Table 6). Agreement was also similar across fracture grades at the vertebral level Fig. 4C). At the patient level, however, the κ increased from lower to higher grades (Fig. 4D).

Discussion

We have recently shown that the presence of vertebral fracture in children diagnosed using the GSQ method correlates with clinically important parameters [17, 18]. Instituting this approach to defining vertebral fracture in pediatric clinical practice requires an understanding of the method's reproducibility in children. In our evaluation of spine radiographs of children with newly diagnosed leukemia, we have found that 4.7% of individual vertebrae were unreadable, with 7.4% unreadable in the Younger group (mean age 3.4 years) vs. 1.6% in the Older group (mean age 9.6 years). These results are similar to findings in another pediatric study that found 7.6% of vertebrae to be unreadable, and is within the spectrum of unreadability reported in the adult literature, which ranges from 0.0% to 12.5% [6, 9, 25–27]. Few data exist on causes of vertebral unreadability on plain radiographs, but vertebral fracture assessment by dual-energy X-ray absorptiometry has a much higher rate of unreadability, which has allowed assessment of some determinants of unreadability. A pediatric vertebral fracture assessment study has reported that visibility is worse in children with lower BMD, a factor that could contribute to lower readability in younger children, who have lower density than older children [26]. In adults, small body size is a cause of impaired readability, suggesting another contributor to lower readability in younger subjects [6, 8, 28–30]. Other possible causes of unreadability reported in adult

vertebral fracture assessment and plain radiographic studies include patient movement, poor radiographic technique, obesity, and anatomical abnormalities of the spine including scoliosis [6, 7, 28]. Intraobserver agreement for determining readability in our study was moderate while interobserver agreement was moderate to substantial

Intraobserver agreement for diagnosis of vertebral fracture was moderate to substantial while interobserver agreement was moderate. The finding of lower interobserver agreement compared to intraobserver agreement is a standard result in most studies of observer variability, since besides the factors that lead to variability by an individual there are additional sources of inter-person inconsistencies (See Table 1). Our diagnostic agreement results are similar to those of a pediatric study using a different set of morphometric definitions that showed intraobserver κ of 0.47 to 1.0 for spine sub-regions and overlap those reported in the adult literature using the GSQ method (Table 1) [1–9, 20]. Some of the causes for diagnostic variability may be the same as those noted for variability in readability, namely small body size and difficulty defining vertebral margins due to lower bone density [6, 26]. This is exemplified by an attempt to use vertebral fracture assessment to image vertebral fractures in a small pediatric cohort [26]. Compared to morphometry on standard spine radiographs as the gold standard, sensitivity by consensus of two observers was only 16% (95% CI, 7–35%), as the computer edge detection paradigm was unable to accurately define vertebral margins [26].

Our analysis included different definitions of fracture for the fracture/non-fracture dichotomy in determining the κ statistic. We found that there was only minimal change at the vertebral level with an increasing fracture grade threshold for either intraobserver or interobserver agreement, and also at the patient level for intraobserver agreement, consistent with minimal change in the weighted κ (weighted by fracture grade) compared to unweighted ones. In contrast, interobserver agreement improved substantially with higher fracture grades at the patient level. The κ rose from fair to moderate at grade 0.5 to substantial or almost perfect at fracture grade 3. This improvement in performance with higher fracture grade has been described in adult morphometry using the GSQ method and is consistent with the view that the most difficult assessment using the GSQ method is in distinguishing a borderline grade 1 fracture from normal [7, 8, 16, 31, 32]. The clinical importance of this finding is that higher-grade fractures confer larger clinical risks [17, 18].

The greatest difficulties with readability and observer agreement on readability were seen in the upper thoracic spine. Although intraobserver agreement on vertebral fracture diagnosis did not vary substantially along the spine, interobserver concordance did relate to anatomical distribution, again with greater variability in the upper thoracic spine. Our results are again similar to findings in adults and to a report in children [1, 3, 6, 28, 29, 32, 33]. These findings have a practical clinical application, as the mid- to upper-thoracic spine (T5 to T8) is one of the major foci of vertebral fractures in children and adolescents [11, 17, 26, 34]. If a vertebral fracture is suspected in the upper thoracic spine of a child, a high standard of radiographic quality is crucial; even then, standard radiographs may be inadequate for visualization and alternate imaging methods may be necessary.

A key limitation of this study is common to many investigations of observer variability: high rates of normal findings and high rates of simple agreement tend to underestimate the true level of concordance [30]. This is exemplified by the disproportionately low reproducibility of fracture diagnosis at L4 (Figure 3), a consequence of the fact that only a single fracture was present at that location in our population compared to at least four fractures at all other vertebrae. If a pair of observations by a single radiologist or between two radiologists did not designate that vertebra as fractured on both occasions, the resulting κ is near zero despite the fact that agreement on non-fracture ranged from 96 to 99%. Some adult studies circumvent this by enriching the study mix with a higher percentage of abnormal findings [1, 2, 5, 8, 29]. This may be one of the reasons the κ values in our study are lower than some of the reported adult values. We chose to utilize our entire study population with ALL so that the calculated levels of agreement reflect real-life performance in this cohort.

Another limitation is the lack of information about normal vertebral height ratios in unfractured vertebral bodies of children. Normal height ratios have been extensively assessed in adults and there is a very low likelihood of any of them exceeding the 20% height ratio reduction threshold used to define fracture [35, 36]. The one anatomical area where there is a low rate of vertebrae exceeding this ratio is the mid-thoracic spine, where some anterior wedging is normally present that produces the natural thoracic kyphosis, leading to some false positives in this region of the spine using the GSQ method [31, 35, 36]. The rate of false positives is sufficiently low that the GSQ threshold of 20% is applied throughout the spine, including the mid-thoracic region, in clinical practice recommendations [13–15]. Solid data on normal height ratios in children is limited to the lower thoracic spine and lumbar area, where values are similar to those of adults and do not exceed 20% [37]. Since we have previously shown that GSQ-method defined fractures in our ALL population correlate with clinical variables like bone density, metacarpal cortical area, back pain, and risk of future fractures, it is likely that the vast majority of the vertebral deformities meeting the GSQ method definition of fracture do, in fact, represent true fractures [17, 18, 38]. Further information is needed on normal vertebral body height ratios in the upper and middle thoracic spine of children and adolescents, however, in order to determine the likelihood of false positives in this region.

A third limitation of our study common to almost all observer variability reports is that only a small number of observers were used, which limits the ability to statistically compare variability between individuals. An additional limitation of our study is that we looked only at prevalent vertebral fractures. Further analyses will be necessary to determine whether similar reproducibility is seen when applying the GSQ method to incident fracture analysis in children. In adults the reproducibility is similar for both prevalent and incident fractures, but this remains to be seen in pediatric imaging [1, 2, 39]. As a final point, our study specifically involved children with ALL. It is conceivable that the underlying disease process altered imaging characteristics of vertebrae and that this contributed to observer variability [40]. It may be that observer variability regarding vertebral fractures would be different in other medical conditions, although this has not been demonstrated to be the case in adults.

Conclusion

Using the GSQ method, intraobserver agreement on vertebral readability was moderate to substantial while interobserver agreement was fair to moderate. Agreement on the presence of vertebral fracture was moderate to substantial at both intra- and interobserver levels, within the range of performance of the technique in adult studies. This technique should be considered for use in pediatric research and clinical practice.

Acknowledgments

The STeroid-associated Osteoporosis in the Pediatric Population (STOPP) Consortium: Alberta Children's Hospital, Calgary, Canada: Reinhard Kloiber, Victor Lewis, Julian Midgley, Paivi Miettinen, David Stephure; British Columbia Women's Hospital and Health Sciences Center, Vancouver, Canada: Brian C. Lentle; British Columbia Children's Hospital, Vancouver, Canada: David Cabral, David B. Dix, Kristin Houghton, Helen R. Nadel; Brock University, St. Catharines, Canada: John Hay; Children's Hospital of Eastern Ontario, Ottawa, Canada: Ciaran Duffy, Janusz Feber, Jacqueline Halton, Roman Jurencak, MaryAnn Matzinger, Johannes Roth, Nazih Shenouda, Leanne M. Ward; London Health Sciences Centre, London, Canada: Elizabeth Cairney, Cheril Clarson, Guido Filler, Joanne Grimmer, Scott McKillop; Keith Sparrow, Robert Stein; IWK Health Center, Halifax, Canada: Elizabeth Cummings, Conrad Fernandez, Adam M. Huber, Bianca Lang, Kathy O'Brien; McMaster Children's Hospital, Hamilton, Canada: Steve Arora, Stephanie Atkinson, Ronald Barr, Craig Coblenz, Peter B. Dent, Maggie Larché, Colin Webber; Montreal Children's Hospital, Montreal, Canada: Sharon Abish, Lorraine Bell, Claire LeBlanc, Celia Rodd, Rosie Scuccimarrì; Ottawa Hospital Research Institute, Ottawa, Canada: David Moher, Tim Ramsay; Shriners Hospital for Children, Montreal, Canada: Francis Glorieux, Frank Rauch; Ste. Justine Hospital, Montreal, Canada: Nathalie Alos, Josee Dubois, Caroline Laverdiere, Veronique Phan, Claire Saint-Cyr; Stollery Children's Hospital, Edmonton, Canada: Robert Couch, Janet Ellsworth, Maury Pinsk, Kerry Siminoski, Beverly Wilson; Université de Sherbrooke, Sherbrooke, Canada: Isabelle Gaboury; Toronto Hospital for Sick Children, Toronto, Canada: Martin Charron, Diane Hebert, Ronald Grant; Winnipeg Children's Hospital, Winnipeg, Canada: Tom Blydt-Hansen, Sara Israels, Kiem Oen, Martin Reed, Shayne Taback.

This study was primarily funded by an operating grant from the Canadian Institutes for Health Research. Additional funding for this work has been provided to Dr. Leanne Ward by the Canadian Institutes for Health Research New Investigator Program, the Canadian Child Health Clinician Scientist Career Enhancement Program and by a University of Ottawa Research Chair Award. The study has also been partially funded by the Children's Hospital of Eastern Ontario and Women and Children's Health Research Institute, University of Alberta.

STOPP would like to thank the following:

The children and their families who participated in the study and without whom the STOPP research program would not have been possible.

The research associates who took care of the patients: Claude Belleville, Ronda Blasco, Erika Bloomfield, Dan Catte, Heather Cosgrove, Valerie Gagne, Diane Laforte, Maritza Laprise, Leila MacBean, Josie MacLennan, Natacha Gaulin Marion, Germaine McInnes, Amanda Mullins, Michele Petrovic, Eileen Pyra, Mala Ramu, Catherine Riddell, Angelyne Sarmiento, Terry Viczko and Aleasha Warner.

The research nurses and support staff from the various Divisions of Nephrology, Oncology, Rheumatology, and Radiology who have contributed to the care of the children enrolled in the study.

The research associates who managed the study at the Children's Hospital of Eastern Ontario: Steve Anderson, Victor Konji, Catherine Riddell, Maya Scharke, Elizabeth Sykes, and Monica Tomiak.

References

1. Genant HK, Wu CY, van Kuijk C, Nevitt MC. Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res.* 1993; 8:137–48.
2. Wu CY, Li J, Jergas M, Genant HK. Comparison of semiquantitative and quantitative techniques for the assessment of prevalent and incident vertebral fractures. *Osteoporos Int.* 1995; 5:354–370. [PubMed: 8800786]

3. Takada M, Wu CY, Lang TF, Genant HK. Vertebral fracture assessment using the lateral scoutview of computed tomography in comparison with radiographs. *Osteoporos Int.* 1998; 8:197–203. [PubMed: 9797902]
4. Wu C, van Kuijk C, Li J, et al. Comparison of digitized images with original radiography for semiquantitative assessment of osteoporotic fractures. *Osteoporos Int.* 2000; 11:25–30. [PubMed: 10663355]
5. Grados F, Roux C, de Vernejoul MC, Utard G, Sebert JL, Fardellone P. Comparison of four morphometric definitions and a semiquantitative consensus reading for assessing prevalent vertebral fractures. *Osteoporos Int.* 2001; 12:716–22. [PubMed: 11605736]
6. Damiano J, Kolta S, Porcher R, Tournoux C, Dougados, Roux C. Diagnosis of vertebral fractures by vertebral fracture assessment. *J Clin Densitom.* 2006; 9:66–71. [PubMed: 16731433]
7. Shousboe JT, DeBold CR. Reliability and accuracy of vertebral fracture assessment with densitometry compared to radiography in clinical practice. *Osteoporos Int.* 2006; 17:281–289. [PubMed: 16172798]
8. Fuerst T, Wu C, Genant HK, et al. Evaluation of vertebral fracture assessment by dual X-ray absorptiometry in a multicenter setting. *Osteoporos Int.* 2009; 20:1199–1205. [PubMed: 19083074]
9. Sanfelix-Genoves J, Arana E, Sanfelix-Gimeno G, Peiro S, Graells-Ferrer M, Vega-Martinez M. Agreement between semi-automatic radiographic morphometry and Genant semi-quantitative method in the assessment of vertebral fracture. *Osteoporos Int.* 2011; 23:2129–2134. [PubMed: 22170523]
10. Cooper C, Dennison EM, Leufkens HG, Bishop N, van Staa TP. Epidemiology of childhood fractures in Britain: a study using the general practice research database. *J Bone Miner Res.* 2004; 19:1976–1981. [PubMed: 15537440]
11. Clark P, Letts M. Trauma to the thoracic and lumbar spine in the adolescent. *Can J Surg.* 2001; 44:337–345. [PubMed: 11603746]
12. Olmez N, Kaya T, Gunaydin R, Dirim B, Erdogan N, Memia A. Intra- and interobserver variability of Kleerekoper's method in vertebral fracture assessment. *Clin Rheumatol.* 2005; 24:215–218. [PubMed: 15940554]
13. Guglielmi G, Diacinti D, van Kuikl C, et al. Vertebral morphometry: current methods and recent advances. *Eur Radiol.* 2008; 18:1484–1496. [PubMed: 18351350]
14. Lentle BC, Brown JP, Khan A, et al. Recognizing and reporting vertebral fractures: reducing the risk of future osteoporotic fractures. *Can Assoc Radiol J.* 2007; 58:27–36. [PubMed: 17408160]
15. Schousboe JT, Vokes T, Broy SB, et al. Vertebral fracture assessment: the 2007 ISCD official positions. *J Clin Densitom.* 2008; 11:92–108. [PubMed: 18442755]
16. Genant HK, Jergas M, Palermo L, et al. Comparison of semiquantitative visual and quantitative morphometric assessment of prevalent and incident vertebral fractures in osteoporosis. *J Bone Miner Res.* 1996; 11:984–96. [PubMed: 8797120]
17. Halton J, Gaboury I, Grant R, et al. 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.* 2009; 24:1326–1334. [PubMed: 19210218]
18. Huber AM, Gaboury I, Cabral DA, et al. Prevalent vertebral fractures among children initiating glucocorticoid therapy for the treatment of rheumatic disorders. *Arthritis Care Res.* 2010; 62:516–528.
19. Ward LM. Osteoporosis due to glucocorticoid use in children with chronic illness. *Horm Res.* 2005; 64:209–221. [PubMed: 16227699]
20. Mäkitie O, Doria AS, Henriques F, et al. Radiographic vertebral morphology: a diagnostic tool in pediatric osteoporosis. *J Pediatr.* 2005; 146:395–401. [PubMed: 15756228]
21. Land C, Rauch F, Munns CF, Sahebjam S, Glorieux FH. Vertebral morphometry in children and adolescents with osteogenesis imperfecta: effect of intravenous pamidronate treatment. *Bone.* 2006; 39:901–906. [PubMed: 16730480]
22. Sumnik Z, Land C, Rieger-Wettengl G, Korber F, Stabrey A, Schoenau E. Effect of pamidronate treatment on vertebral deformity in children with primary osteoporosis. *Horm Res.* 2004; 61:137–142. [PubMed: 14676462]

23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–74. [PubMed: 843571]
24. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1968; 70:213–220. [PubMed: 19673146]
25. Rea JA, Li J, Blake GM, Steiger P, Genant HK, Fogelman I. Visual assessment of vertebral deformity by X-ray absorptiometry: a highly predictive method to exclude vertebral deformity. *Osteoporos Int*. 2000; 11:660–668. [PubMed: 11095168]
26. Mayranpaa MK, Helenius I, Valtia H, Mayranpaa MI, Toivianen-Salo S, Makitie O. Bone densitometry in the diagnosis of vertebral fractures in children: accuracy of vertebral fracture assessment. *Bone*. 2007; 41:353–359. [PubMed: 17618848]
27. Rea JA, Chen MB, Li J, et al. Morphometric x-ray absorptiometry and morphometric radiography of the spine: a comparison of prevalent vertebral deformity identification. *J Bone Miner Res*. 2000; 15:564–574. [PubMed: 10750572]
28. Chapurlat RD, Duboeuf F, Marion-Audibert HO, Kalpakcioglu B, Mitlak BH, Delmas PD. Effectiveness of instant vertebral assessment to detect prevalent vertebral fracture. *Osteoporos Int*. 2006; 17:1189–1195. [PubMed: 16758133]
29. Hoppers IC, van der Laan JG, Zeebregts CJ, et al. Vertebral fracture assessment in supine position: comparison by using conventional semiquantitative radiography and visual radiography. *Radiology*. 2009; 251:822–828. [PubMed: 19380691]
30. Gwet K. Inter-rater reliability: dependency on trait prevalence and marginal homogeneity. *Statistical Methods for Inter-rater Reliability Assessment*. 2002; 2:1–9.
31. Black DM, Palermo L, Nevitt MC, et al. Comparison of methods for defining prevalent vertebral deformities: the Study of Osteoporotic Fractures. *J Bone Miner Res*. 1995; 0:890–902.
32. Fechtenbaum J, Cropet C, Kolta S, Verdoncq B, Orcel P, Roux C. Reporting of vertebral fracture on spine x-rays. *Osteoporos Int*. 2005; 16:1823–1826. [PubMed: 16032373]
33. Pearson D, Horton B, Green DJ, Hosking DJ, Goodby A, Steel SA. Vertebral morphometry by DXA: a comparison of supine lateral and decubitus lateral densitometers. *J Clin Densitom*. 2006; 9:295–301. [PubMed: 16931347]
34. Siminoski K, Lee K-C, Jen H, et al. Anatomical distribution of vertebral fractures: comparison of pediatric and adult spines. *Osteoporos Int*. 2011; 23:1999–2008. [PubMed: 22109742]
35. Szulc P, Munoz F, Sornay-Rendu E, et al. Comparison of morphometric assessment of prevalent vertebral deformities in women using different reference data. *Bone*. 2000; 27:841–846. [PubMed: 11113396]
36. Jackson SA, Tenenhouse A, Robertson L. Vertebral fracture definition from population-based data: preliminary results from the Canadian Multicenter Osteoporosis Study (CaMos). *Osteoporos Int*. 2000; 11:680–687. [PubMed: 11095171]
37. Gaca AM, Barnhart HX, Bisset GS. Evaluation of wedging of lower thoracic and upper lumbar vertebral bodies in the pediatric population. *Am J Roentgenol*. 2010; 194:516–520. [PubMed: 20093618]
38. Alos N, Grant RM, Ramsay T, et al. High incidence of vertebral fractures in children with acute lymphoblastic leukemia 12 months after the initiation of therapy. *J Clin Oncol*. 2012; 30:2760–2767. [PubMed: 22734031]
39. Grados F, Fechtenbaum J, Flipon E, Kolta S, Roux C, Fardellone P. Radiographic methods for evaluating osteoporotic vertebral fractures. *Joint Bone Spine*. 2009; 76:241–247. [PubMed: 19196531]
40. Hanrahan CJ, Shah LM. MRI of spinal bone marrow: Part 2, T1-weighted imaging-based differential diagnosis. *Am J Roentgen*. 2011; 197:1309–1321.

Fig. 1.

Application of the Genant semi-quantitative method. Loss of vertebral body height is visually estimated as a ratio of a reference vertebral height for each of three locations on a vertebral body: anterior, middle, and posterior. The loss of anterior vertebral body height (a), designated X , is assessed in relation to the posterior height of the same vertebral body, designated H . Loss of middle vertebral body height (b) is similarly evaluated in comparison the posterior vertebral height of the same vertebral body. Loss of posterior vertebral body height (c) is assessed by comparison to the posterior vertebral body of the vertebrae above (H_{upper}) and below (H_{lower}). In the case of T4 and L4, only one adjacent vertebra is available, as the complete assessment is from T4 to L4.

Fig. 2.
Mean percentage of vertebrae at each vertebral level that were considered unreadable.

Fig. 3.

Reproducibility of fracture diagnosis at each vertebral level for (a) intraobserver agreement and (b) interobserver agreement

Fig. 4.

Observer agreement according to Genant semi-quantitative method grade used to define fracture for (a) intraobserver agreement at the vertebral level, (b) intraobserver agreement at the patient level, (c) interobserver agreement at the vertebral level, and (d) interobserver agreement at the patient level.

Table 1

Published observer agreement on vertebral fracture diagnosis in adults using the Genant semi-quantitative method [1–9]

First Author	Reference Number	Number of Readers	Intraobserver K	Interobserver K
Vertebral Level				
Genant ^a	1	2	0.73–0.89	0.74
Wu ^a	2	3	N/A	0.80–0.81
Takada ^a	3	2	0.82–1.00 ^d	0.76–0.91 ^d
Wu ^a	4	2	0.87–0.96	N/A
Grados ^b	5	3	0.91	N/A
Damiano ^b	6	2	N/A	0.86
Schousboe	7	2	N/A	0.60
Fuerst ^a	8	3	0.81–0.85	0.75–0.82
Sanfelix-Genoves ^c	9	2	N/A	0.39
Patient Level				
Damiano ^b	6	2	N/A	0.79
Schousboe	7	2	N/A	0.53
Fuerst ^a	8	3	0.73–0.83	0.71–0.81
Sanfelix-Genoves ^c	9	2	N/A	0.26

N/A not available, K Cohen's kappa

^aStudies from same group

^bStudies from same group

^cCalculated from data in paper

^dFor spine segments, not entire spine from T4 to L4

Table 2

Subject characteristics

Parameter	Interobserver Cohort	Interobserver Younger Cohort	Interobserver Older Cohort	Intraobserver Cohort	Intraobserver Younger Cohort	Intraobserver Older Cohort
Number	186	93	93	100	50	50
Male, n (%)	109 (57)	56 (60)	53 (57)	57 (57)	30 (60)	27 (54)
Age (yrs), median (minimum, maximum)	5.3 (1.3, 17.0)	3.4 (1.3, 5.2)*	9.6 (5.3, 17.0)	5.4 (1.3, 16.6)	3.4 (1.3, 5.3)***	10.2 (5.4, 16.6)
White ethnicity, n (%)	140 (75)	74 (80)	66 (71)	76 (76)	41 (82)	35 (70)
Height Z-score, mean (SD)	0.3 (1.1)	0.4 (1.1)	0.1 (1.2)	0.2 (1.2)	0.4 (1.2)	0.0 (1.2)
Weight Z-score, mean (SD)	0.4 (1.2)	0.7 (1.2)**	0.1 (1.2)	0.3 (1.2)	0.6 (1.2)****	-0.1 (1.1)
Lumbar spine BMD ^a Z-score, mean (SD)	-1.2 (1.3)	-1.4 (1.2)	-1.1 (1.4)	-1.3 (1.4)	-1.3 (1.3)	-1.3 (1.5)
Spine BMD Z-score for bone age, mean (SD)	-1.2 (1.3)	-1.3 (1.2)	-1.1 (1.3)	-1.3 (1.4)	-1.3 (1.4)	-1.3 (1.5)
Vertebral fractures per patient, n (%)						
0	157 (84)	81 (87)	76 (82)	79 (79)	43 (86)	36 (72)
1	18 (10)	7 (8)	11 (12)	12 (12)	3 (6)	9 (18)
2 to 5	7 (4)	3 (3)	4 (4)	7 (7)	3 (6)	4 (8)
6 to 10	4 (2)	2 (2)	2 (2)	2 (2)	1 (2)	1 (2)
Number of children defined as worst grade of vertebral fracture, n (%)						

Parameter	Interobserver Cohort	Interobserver Younger Cohort	Interobserver Older Cohort	Intraobserver Cohort	Intraobserver Younger Cohort	Intraobserver Older Cohort
grade 1	14 (48)	7 (50)	7 (41)	11 (52)	4 (57)	7 (50)
grade 2	10 (35)	4 (40)	6 (35)	5 (24)	2 (29)	3 (21)
grade 3	5 (17)	1 (8)	4 (24)	5 (24)	1 (14)	4 (29)

BMD bone mineral density, SD standard deviation.

*
p < 0.001 compared to Interobserver Older Cohort,
**
p = 0.001 compared to Interobserver Older Cohort,

p < 0.001 compared to Intraobserver Older Cohort,

p = 0.004 compared to Intraobserver Older Cohort

Table 3

Observer agreement on vertebral readability

Intraobserver Agreement		Radiologist 1	Radiologist 2	Radiologist 3
Vertebral Level	Agreement (95 % CI)	97 (96, 98)	98 (97, 99)	94 (93, 95)
	κ (95% CI)	0.434 (0.270, 0.599)	0.648 (0.514, 0.782)	0.429 (0.305, 0.554)
Patient Level	Agreement (95% CI)	84 (77, 91)	88 (82, 94)	79 (71, 87)
	κ (95% CI)	0.416 (0.153, 0.678)	0.611 (0.405, 0.818)	0.496 (0.305, 0.688)
Interobserver Agreement		Radiologist 1 vs. 2	Radiologist 1 vs. 3	Radiologist 2 vs. 3
Vertebral Level	Agreement (95 % CI)	97 (96, 97)	93 (92, 94)	94 (94, 95)
	κ (95% CI)	0.398 (0.270, 0.526)	0.330 (0.232, 0.429)	0.504 (0.422, 0.585)
Patient Level	Agreement (95% CI)	82 (76, 87)	70 (63, 76)	76 (70, 82)
	κ (95% CI)	0.351 (0.153, 0.548)	0.295 (0.141, 0.450)	0.467 (0.329, 0.605)

CI confidence interval, *K* Cohen's kappa, *vertebral level* indicates determination of whether individual vertebrae were unreadable, *patient level* indicates determination of whether a patient had one or more unreadable vertebral bodies

Table 4

Observer agreement on vertebral fracture diagnosis

Intraobserver Agreement				
		Radiologist 1	Radiologist 2	Radiologist 3
	Agreement (95 % CI)	98 (97, 99)	97 (96, 98)	96 (95, 97)
Vertebral Level	κ (95% CI)	0.726 (0.620, 0.833)	0.668 (0.566, 0.771)	0.529 (0.402, 0.655)
	Weighted κ	0.731 (0.637, 0.826)	0.678 (0.597, 0.758)	0.632 (0.531, 0.733)
	Agreement (95% CI)	92 (87, 97)	89 (83, 95)	84 (77, 91)
Patient Level	κ (95% CI)	0.767 (0.612, 0.922)	0.642 (0.443, 0.842)	0.528 (0.316, 0.740)
	Weighted κ	0.766 (0.637, 0.895)	0.778 (0.639, 0.917)	0.678 (0.515, 0.841)
Interobserver Agreement				
		Radiologist 1 vs. 2	Radiologist 1 vs. 3	Radiologist 2 vs. 3
	Agreement (95 % CI)	97 (96, 97)	96 (95, 97)	96 (95, 96)
Vertebral Level	κ^a (95% CI)	0.532 (0.430, 0.633)	0.455 (0.348, 0.562)	0.548 (0.461, 0.635)
	Weighted κ	0.531 (0.452, 0.610)	0.481 (0.393, 0.570)	0.568 (0.496, 0.641)
	Agreement (95% CI)	85 (80, 90)	82 (76, 87)	82 (76, 87)
Patient Level	κ (95% CI)	0.486 (0.310, 0.661)	0.433 (0.261, 0.605)	0.470 (0.308, 0.631)
	Weighted κ	0.573 (0.434, 0.713)	0.525 (0.384, 0.666)	0.592 (0.457, 0.728)

CI confidence interval, κ Cohen's kappa, *vertebral level* indicates determination of whether individual vertebrae were fractured, *patient level* indicates determination of whether a patient had one or more vertebral fractures

Table 5

Intraobserver agreement on vertebral fracture diagnosis for Younger and Older age groups

		Radiologist 1	Radiologist 2	Radiologist 3
Younger Group				
Vertebral Level	Agreement (95 % CI)	98 (97, 99)	99 (98, 100)	96 (95, 98)
	κ (95% CI)	0.656 (0.464, 0.849)	0.828 (0.692, 0.965)	0.503 (0.300, 0.707)
Patient Level	Agreement (95% CI)	94 (87, 100)	98 (94, 100)	84 (74, 94)
	κ (95% CI)	0.805 (0.591, 1.000)	0.878 (0.641, 1.000)	0.412 (0.038, 0.785)
Older Group				
Vertebral Level	Agreement (95 % CI)	98 (97, 99)	95 (93, 97)	95 (94, 97)
	κ (95% CI)	0.769 (0.645, 0.893)	0.597 (0.463, 0.731)	0.546 (0.384, 0.707)
Patient Level	Agreement (95% CI)	90 (82, 98)	80 (69, 91)	84 (74, 94)
	κ (95% CI)	0.733 (0.512, 0.955)	0.527 (0.265, 0.789)	0.598 (0.342, 0.853)

CI confidence interval, *K* Cohen's kappa, *vertebral level* indicates determination of whether individual vertebrae were unreadable, *patient level* indicates determination of whether a patient had one or more unreadable vertebral bodies, Younger group had mean age 3.4 years (range 1.3 to 5.3 years), Older group had mean age 10.2 years (range 5.4 to 16.6 years)

Table 6

Interobserver agreement on vertebral fracture diagnosis for Younger and Older age groups

		Radiologist 1 vs. 2	Radiologist 1 vs. 3	Radiologist 2 vs. 3
Younger Group				
Vertebral Level	Agreement (95 % CI)	98 (97, 99)	96 (95, 97)	97 (96, 98)
	κ (95% CI)	0.606 (0.456, 0.755)	0.488 (0.331, 0.646)	0.613 (0.483, 0.743)
Patient Level	Agreement (95% CI)	88 (82, 95)	84 (76, 91)	87 (80, 94)
	κ (95% CI)	0.524 (0.259, 0.788)	0.422 (0.155, 0.690)	0.549 (0.310, 0.787)
Older Group				
Vertebral Level	Agreement (95 % CI)	96 (94, 97)	95 (94, 96)	94 (93, 96)
	κ (95% CI)	0.484 (0.348, 0.620)	0.429 (0.283, 0.575)	0.505 (0.389, 0.621)
Patient Level	Agreement (95% CI)	82 (74, 90)	80 (71, 88)	76 (68, 85)
	κ (95% CI)	0.453 (0.218, 0.688)	0.436 (0.210, 0.662)	0.400 (0.180, 0.619)

CI confidence interval, *K* Cohen's kappa, *vertebral level* indicates determination of whether individual vertebrae were unreadable, *patient level* indicates determination of whether a patient had one or more unreadable vertebral bodies, Younger group had mean age 3.4 years (range 1.3 to 5.2 years), Older group had mean age 9.6 years (range 5.3 to 17.0 years)