

Value and potential limitations of vertebral fracture assessment (VFA) compared to conventional spine radiography: experience from a fracture liaison service (FLS) and a meta-analysis

F. Malgo¹  · N. A. T. Hamdy¹ · C. H. J. M. Ticheler² · F. Smit² · H. M. Kroon² · T. J. Rabelink¹ · O. M. Dekkers^{3,4} · N. M. Appelman-Dijkstra¹

Received: 31 December 2016 / Accepted: 21 June 2017 / Published online: 25 August 2017
© The Author(s) 2017. This article is an open access publication

Abstract

Summary We evaluated the value of VFA in the identification of vertebral fractures using a retrospective study and a meta-analysis. Performance of VFA was adequate in the meta-analysis although this was not demonstrated in our centre. We recommend checking the performance of VFA tools before exclusively relying on this tool.

Introduction Vertebral fractures are traditionally diagnosed using conventional radiographs of the spine. Vertebral fracture assessment (VFA) has been advocated as an alternative tool in the diagnosis of these fractures.

Methods We conducted a retrospective study as well as a systematic review and a meta-analysis to evaluate the performance of VFA compared to conventional spinal radiography in patients who had sustained a fracture and thus at risk for osteoporosis. A risk of bias analysis was also performed.

Results The diagnostic study included 542 patients (25% male) with fractures. The sensitivity of low-radiation VFA to detect a patient with a vertebral fracture \geq Genant grade 2 was

0.77 and its specificity 0.80. Two hundred ninety-seven (55%) patients had ≥ 1 and 135 (25%) ≥ 3 unevaluable vertebrae. The systematic review identified 16 studies including a total of 3238 subjects (19% male) with a mean age range of 45 to 74 years. Seven studies had a low risk of bias and 9 had an intermediate risk, mainly due to not consecutively including patients. The pooled sensitivity of VFA to detect a patient with a vertebral fracture \geq Genant grade 2 was 0.84 (95% CI, 0.72–0.92) and specificity 0.90 (95% CI, 0.84–0.94).

Conclusions Our findings from the meta-analysis suggest an adequate performance of VFA for the detection of vertebral fractures. However, we could not demonstrate these findings in our center, especially the specificity. Our data advocate caution with exclusively relying on VFA in the assessment of vertebral fractures without identifying performance and potential limitations of the technique.

Keywords Sensitivity · Spinal radiography · Vertebral fracture · Vertebral fracture assessment · X-rays

Electronic supplementary material The online version of this article (doi:10.1007/s00198-017-4137-6) contains supplementary material, which is available to authorized users.

✉ N. M. Appelman-Dijkstra
N.M.Appelman-Dijkstra@lumc.nl

¹ Department of Medicine, Division Endocrinology and Center for Bone Quality, C7-Q, Leiden University Medical Center, P.O. Box 9600, 2300 RC Leiden, The Netherlands

² Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands

³ Department of Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

⁴ Department of Clinical Epidemiology, Institute of Clinical Medicine, Aarhus University Hospital, Aarhus, Denmark

Introduction

Vertebral fractures are associated with increased mortality and morbidity and decreased quality of life, and the incidence of these fractures increases with age [1–4]. The prevalence and grade of severity of vertebral fractures have also been shown to be predictive for the risk of new vertebral and non-vertebral fractures, independently of bone mineral density (BMD) measurements [5–7]. However, vertebral fractures remain often underdiagnosed despite their clear value in the assessment of fracture risk [8, 9]. Conventional spine radiography is traditionally used in the evaluation of vertebral fractures and is considered to be the “gold standard” for detection of these fractures and their grading using the semi-quantitative method of Genant [10]. Vertebral fracture

assessment (VFA) is performed using images obtained by bone densitometers in the same session as bone mineral density measurements are performed to screen for osteoporosis. The VFA technique enables the acquisition of a patient-friendly alternative to conventional radiographs for the assessment of vertebral fractures in a one-stop diagnostic test [11]. Other advantages of VFA include lower radiation exposure and possibly lower costs. On the basis of available data, VFA has indeed already been incorporated in a number of clinical guidelines replacing conventional radiography for the assessment of prevalent vertebral fractures and thus for the risk of fracture [12, 13].

However, the advantage of lower-radiation doses used in certain bone densitometry scanners are associated with the drawback of poor image quality and thus of potential poor visualization of the contours of the vertebrae, which could lead to misclassification of fractures or the ascertaining of a vertebra as non-evaluable leading to an inaccurate estimation of fracture risk. A standard protocol or technique for performing VFA has never actually been developed, and the majority of published studies compare the performance of VFA to that of conventional spine radiographs mostly in diverse patient populations, often consisting of small numbers and mostly using different hardware and radiation protocols.

The aims of our study were twofold: first to evaluate the performance of VFA compared to conventional spine radiography in our fracture liaison service (FLS) to assess whether we could replace conventional radiographs by VFA in the diagnosis of vertebral fractures in patients evaluated for osteoporosis after a recent fracture. Second, to systematically review all published literature on the performance of VFA compared to conventional spine radiography in patients evaluated for suspected osteoporosis and to perform a meta-analysis on these data.

Methods

Vertebral fracture assessment: VFA compared to conventional radiography

Study design

This was a retrospective study evaluating the performance of low-radiation single energy x-ray absorptiometry VFA for the detection of vertebral fractures compared to conventional radiography of the spine, in a cohort of consecutive men and women aged 50 years or older who had sustained a fracture between June 2012 and June 2014 and who were assessed for osteoporosis according to screening protocols used in the FLS of the Leiden University Medical Center [14]. In these protocols, all patients attending the FLS are screened, diagnosed and treated for osteoporosis where required and data collected at source in a database. Because of the nature of the study, the

Medical Ethics Committee of the Leiden University Medical Center deemed that no written informed consent was required.

For the purpose of this analysis, only patients with available data on both VFA and conventional radiography were included in the study. The following data were retrieved from the database: age, gender, height, weight, a detailed fracture history, family history of osteoporosis, a list of current medication and history of use of bone modifying agents were obtained.

Bone mineral density measurements

Bone mineral density was measured at the lumbar spine (L1–L4) and at the left and right femoral neck by dual-energy X-ray absorptiometry (DXA) using Hologic QDR 4500 (Hologic, Bedford, MA, USA). NHANES III reference values compatible with reference values for the Dutch population were used to calculate T scores. The diagnosis of osteoporosis, osteopenia or normal BMD was established using the World Health Organization criteria.

Vertebral fracture assessment

In addition to the BMD measurements, single-energy x-ray lateral VFA images of the spine (T4–L4) were obtained by a dedicated technician with the patient lying in supine position and a cushion supporting the knees. The effective radiation dose of a VFA scan received by the patient is typically 3 microSievert.

Conventional radiography of the spine

Antero-posterior (thoracic spine), postero-anterior (lumbar spine) and lateral conventional radiographs of the thoracic and lumbar spine were performed by a radiology technician using a standardized protocol, with the detector centralized on Th7 for the thoracic spine and on L3 for the lumbar spine.

Assessment of vertebral fractures using VFA and conventional radiography of the spine

The presence of vertebral fractures was assessed on VFA images using Hologic QDR Physician Viewer software. The software generates six points on each vertebral endplate which were then manually adjusted by a dedicated technician as required. In the majority of vertebrae (more than 90%), manual adjustment of the automatically placed points had to be performed by a trained laboratory technician. Anterior, middle and posterior corporal heights were calculated automatically. Following this quantitative evaluation, the software further used the criteria for the classification of vertebral fractures as described by Genant [10]. Analyses were performed on a per-person basis. A vertebral fracture was defined as Genant grade 2 or more. All routinely generated reports of conventional radiographs performed as part of the protocol used in the

FLS were retrieved from the patient's electronic medical records. In addition, one of the authors (F.M.) further assessed all radiographs for the presence and grading of vertebral fractures. Both observers were blinded to the VFA findings. Vertebral fractures were classified according to Genant grading system: grade 1 for an anterior, mid or posterior reduction of 20–25% in vertebral height; grade 2 for a reduction of 25–40% and grade 3 for a reduction of more than 40% in vertebral height. In case of disagreement between radiology reports and evaluation from (F.M.), spine radiographs were evaluated by an experienced musculoskeletal radiologist (H.K.), whose evaluation was decisive and used in the analysis. In addition, a randomly selected sample of 20% of the remaining patients was also evaluated by (H.K.) in order to validate the classification of vertebral fractures based on the combined report of the FLS charts and (F.M.), which yielded a kappa of 0.82.

Systematic review of literature

Search strategy

We designed a search strategy in collaboration with a trained librarian for studies that primarily focussed on the diagnostic accuracy of VFA compared to conventional radiographs of the spine in the diagnosis and grading of vertebral fractures in patients at risk for osteoporosis. The search was conducted in PubMed, MEDLINE, EMBASE and Web of Science and included all published articles on the topic up to June 10, 2016. All relevant keywords were used, including free text words. The complete search strategy is provided as Supplemental Data.

Eligibility criteria and data extraction

Only original articles written in English were included. Inclusion criteria were (1) comparison between VFA and spine radiographs performed for the diagnosis of vertebral fractures with reported data on sensitivity and specificity, (2) suspicion of osteoporosis as indication for the assessment of vertebral fractures, (3) use of the Genant's or ABQ's method to assess the presence of vertebral fractures in radiography and (4) patients aged ≥ 18 years. Studies that reported patients with diseases of the spine such as ankylosing spondylitis or patients recruited from the general population were not eligible.

Articles were assessed by two independent investigators (F.M. and N.M.A-D), first by screening for eligibility for inclusion in the analysis by title and abstract. Selected articles were further assessed in detail. Disagreements were resolved by consensus.

The following data were extracted from all selected publications: number of patients studied, age and gender distribution, hardware used for VFA and DXA, study inclusion criteria,

method of assessment of vertebral fractures, prevalence of vertebral fractures and sensitivity and specificity of VFA.

Risk of bias assessment

The following characteristics of the study design were evaluated for each published study used in the review to assess the risk of bias:

1. Inclusion of patients; were consecutive patients who had conventional spine radiographs included in the study or were only selected patients included? Inclusion of consecutive patients was considered a low risk of bias.
2. Definition of vertebral fractures used in the study. Analysis of data using a definition of Genant 2 or higher for vertebral fractures was considered a low risk of bias [10].
3. Clear and adequate description of method used to assess vertebral fractures in VFA and conventional radiography of the spine. Complete description of methodology for the assessment of vertebral fractures was considered a low-risk of bias.
4. Blinding of the examiner who examined VFA for the outcome of the spine radiographs. Blinded assessment was considered a low risk of bias.

For each of the four elements named above, studies were qualified as adequate, inadequate or not reported.

Statistical analysis

The performance of VFA was calculated using conventional radiography as reference, sensitivity was estimated by the number of true-positive vertebral fractures divided by the number of vertebral fractures identified by conventional radiographs, and specificity was calculated by the number of true-negative vertebral fractures divided by the number of intact vertebrae observed on conventional radiographs. The main outcome of the meta-analysis was the pooled sensitivity and specificity of VFA. Conventional radiographs of the spine were used as the gold standard. The meta-analysis was based on a random effects model and a bivariate approach, and sensitivity and specificity were estimated both per vertebra and per person. Heterogeneity was assumed and explored as recommended by Leeflang et al. [15].

Results

Vertebral fracture assessment: VFA compared to conventional radiography

Five hundred and forty-two patients [137 (25%) men and 405 (75%) women] were included in the study. Mean age of the

population was 67.5 ± 10.1 years (range 50.0–92.8), mean BMI was 26.1 ± 4.3 kg/m² and median time between fracture and FLS visit was 2.3 months. Fifty patients (9%) had sustained a fracture of the hip, 25 (5%) of the vertebrae, 188 (35%) of the distal radius, 58 (11%) of the proximal humerus and 61 (11%) of the ankle. The majority of patients had osteopenia ($n = 319$, 59%), 163 (30%) had osteoporosis and 60 (11%) had a normal bone mineral density (Table 1). On low-radiation VFA, 184 (34%) patients had at least one grade 2 or higher vertebral fracture, of which 47 had a Genant grade 3 vertebral fracture. These were 56 men and 128 women with a mean age of 71.4 ± 10.3 years. One hundred and six (58%) patients had osteopenia, 64 (35%) osteoporosis and 14 (8%) had a normal BMD.

Conventional radiographs of the spine identified 132 (24%) patients with \geq one grade 2 or higher vertebral fracture, of which 47 had a Genant grade 3 vertebral fracture. VFA correctly identified 102 of the 132 patients with a \geq grade 2 vertebral fracture, corresponding to a sensitivity of 0.77 (95% CI, 0.70–0.84). Of the 30 patients who were missed on VFA, 17 had ≥ 1 vertebrae that could not be evaluated by VFA. Of these patients, three had a radiological fracture on their radiographs at a vertebral level that could not be evaluated by VFA, and were thus missed. Of the 410 patients without a vertebral fracture on spinal radiographs, 328 were also found not to have a vertebral fracture on VFA corresponding to a specificity of 0.80 (0.76–0.84) (Table 2). Interestingly, 297 (55%) patients had ≥ 1 vertebrae that could not be evaluated by VFA and 135 (25%) patients had 3 or more unevaluable vertebrae (Fig. 1).

The vertebrae that could not be evaluated by VFA or the patients that were misclassified by VFA were independent of the type of the recently sustained fracture or of the time between the recent fracture and the FLS visit.

Search strategy

The search strategy for the systematic review of the literature yielded 694 articles (201 from PubMed, 167 from MEDLINE, 203 from Embase and 123 from Web of Science). Two hundred seventy studies were unique and potentially relevant and were further assessed for eligibility. Two hundred forty-one studies were excluded on the basis of title and abstract, 14 were review papers and 1 was a Position Paper; 2 studies were performed in a paediatric population, 1 study was written in French and 1 study could not be obtained. Twenty-nine studies were acquired for full assessment. Of these, 4 were excluded because there was no comparison between VFA and conventional radiography, 2 studies did not report performance parameters, 3 studies included patients with a rheumatologic disorder, 1 study included patients from the general population and 1 study was an autopsy study (Fig. 2).

Table 1 Patient characteristics

Age (years)	67.5 \pm 10.1
Male/female	137/405
BMI (kg/m ²)	26.1 \pm 4.3
Time between fracture and VFA (months)	2.3 \pm 0.08
Type of fracture ^a	
Vertebra	25 (5%)
Hip	50 (9%)
Proximal humerus	58 (11%)
Ankle	61 (11%)
Distal radius	188 (35%)
DXA-BMD	
Normal	60 (11%)
Osteopenia	319 (59%)
Osteoporosis	163 (30%)

Values are expressed as mean \pm SD, except for time between fracture and VFA

BMI body mass index, kg/m² kilogramme per square meter, VFA vertebral fracture assessment, DXA-BMD bone mineral density measurement

^a Fracture for which the patients was referred to FLS

Eighteen articles met all specified inclusion criteria, two of which reported on related study populations. A total of 16 studies were thus included in the final analysis. Two of these 16 studies included two different populations, namely patients at high and low risk for osteoporosis and/or fractures [16, 17]. In keeping with our inclusion criteria, patients recruited from the general population and thus at low risk for osteoporosis were excluded from the analysis ($n = 582$).

Study characteristics

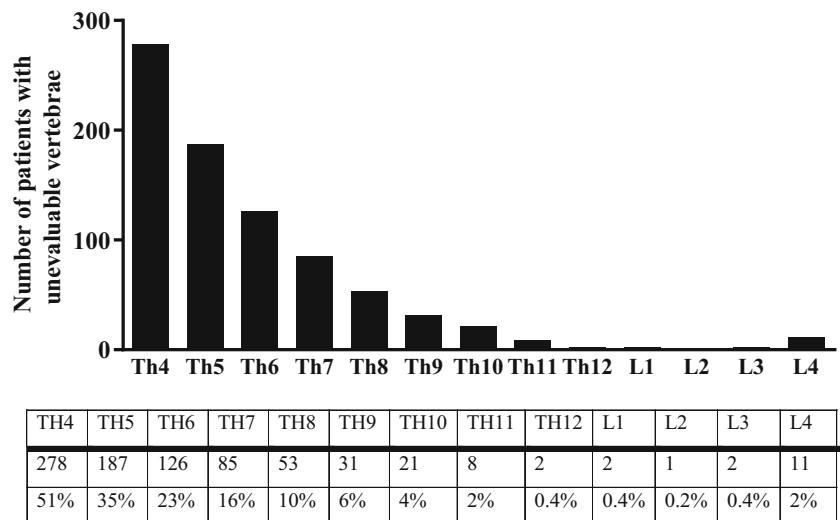
A total number of 3238 subjects were included in the analysis, the vast majority of whom were women ($n = 2626$). The number of subjects per study ranged from 35 [18] to 930 [19] subjects. Mean age of the studied populations ranged from 45 to 74 years. The youngest included patient was 23 years old [20] and the oldest 96 years old [21]. There were seven studies that included both female and male subjects [18, 20, 22–26]. One study included both male and female subjects but did not specify the exact gender distribution of subjects who

Table 2 Outcome of VFA compared to conventional spine radiography for the detection of vertebral fractures \geq grade 2

VFA	Radiography of the spine	
	Vertebral fracture	No vertebral fracture
Vertebral fracture	102	82
No vertebral fracture	30	328
Total	132	410

VFA vertebral fracture assessment

Fig. 1 Number (%) of patients with vertebrae that could not be evaluated by VFA



had conventional spinal radiography in addition to VFA [25] (Table 3).

All studies included subjects recruited from outpatient clinics, and two studies additionally included patients admitted with a recent vertebral fracture [27] or hip fracture [17]. One study solely included patients with radiological evidence of osteoporotic vertebral fractures [16]. Two studies used data on VFA and conventional radiography originally collected for another study [23, 28], one used data from osteoporosis treatment studies and one from an HIV-related osteoporosis study. Three studies reported the inclusion of patients who had recently sustained a fracture [17, 24, 27].

Twelve of the 16 studies used Hologic hardware and five used GE Lunar hardware to acquire VFA scans, with one of the 16 studies acquiring VFA images with either Hologic or GE Lunar technology [29].

Prevalence of vertebral fractures on conventional spine radiography

The prevalence of vertebral fractures \geq grade 1 ranged from 1.8 [22] to 39% [18]; the prevalence of patients with a vertebral fracture \geq grade 1 ranged from 6.9 [30] to 100% [16].

Fig. 2 Flowchart of selection of articles for systematic review and meta-analysis. VFA vertebral fracture assessment

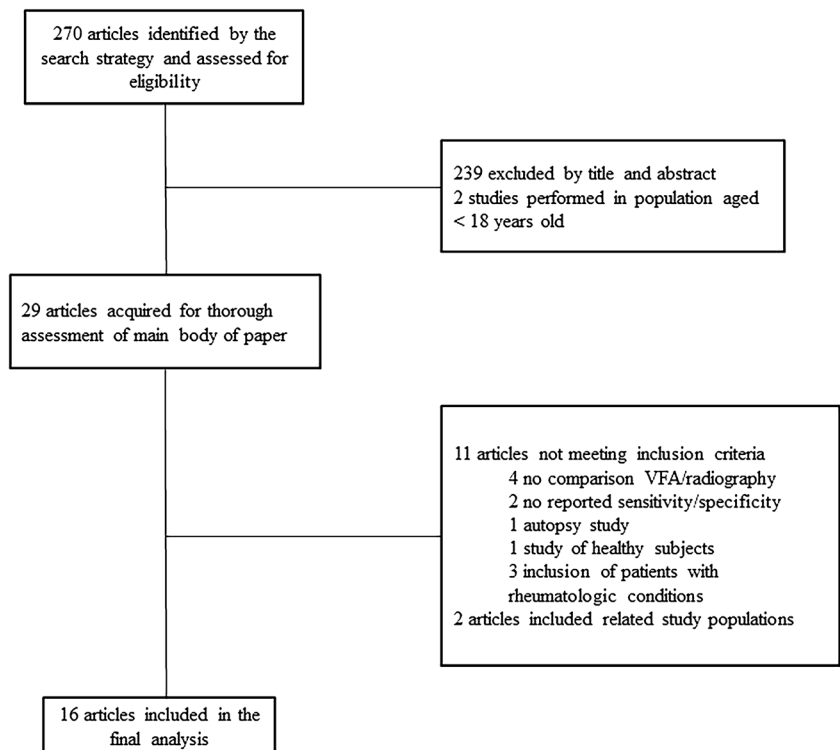


Table 3 Characteristics of studies included in meta-analysis

Author, publication year	Number of patients (m/f)	Age (range)	Hardware	Inclusion criteria	Analysis method	VF prevalence (%)	Outcome parameters (sensitivity/specificity)
Bazzocchi (2012)	68 (38/30)	58.1 (32–83)	Lunar iDXA	Suspicion of osteoporosis, glucocorticoid treatment, organ transplantation)	Genant	Vertebra ≥ grade 1 4.9 ≥ grade 2 1.8 Person ≥ grade 1 38.2	Vertebra ≥ grade 1 70.0 98.3 ≥ grade 2 53.8 99.6 Person ≥ grade 1 73.1 90.5
Binkley (2005)	80 (0/80)	72.8 (61–84)	Lunar Prodigy	PM women participated in osteoporosis treatment studies or having clinical BMD measurement	Genant	Vertebra ≥ grade 1 4.8 ≥ grade 2 2.2 Person	Vertebra ≥ grade 1 70.96 ≥ grade 2 94.99 Person
Chapurlat (2006)	85 (0/85)	71 (NR)	Hologic Delphi	PM women with screening BMD measurement or VF on radiography	Genant	Vertebra ≥ grade 1 9.9 Person ≥ grade 1 50	Vertebra ≥ grade 1 NR NR Person ≥ grade 1 69.74
Damiano (2006)	136 (0/136)	69.1 (37–96)	Hologic Delphi W	PM women with indication for spinal radiography in rheumatology department	Genant	Vertebra ≥ grade 1 9 ≥ grade 2 6 Person ≥ grade 1 45.8	Vertebra ≥ grade 1 82.8 98.3 ≥ grade 2 82.8 99.3 Person ≥ grade 1 94.1 83.0
Deleskog (2016)	35 (5/30)	67.5 (NR)	Hologic Discovery A QDR	Patients referred for Forsteo treatment	Genant	Vertebra ≥ grade 1 39 ≥ grade 2 29 Person	Vertebra ≥ grade 1 76.87 ≥ grade 2 72.92 Person
Diacinti (2012)	350 (81/269)	60.6 (28–85)	Lunar iDXA	Patients referred by GP for osteoporosis and from a study of HIV-related osteoporosis	Genant	Vertebra ≥ grade 1 5.1 Person ≥ grade 1 36	Vertebra ≥ grade 1 97.0 99.9 Person ≥ grade 1 96.8 98.7
Diacinti (2012)	930 (0/930)	62.4 (46–85)	Hologic QDR-4500A	PM women referred by GP for evaluation of osteoporosis	Genant	Vertebra ≥ grade 1 3.7 Person ≥ grade 1 27 Ref: Vertebra ≥ grade 1 1.6 Person	Vertebra ≥ grade 1 97.9 99.8 Person ≥ grade 1 97.2 98.9 Ref: Vertebra ≥ grade 1 54.99 Person
Ferrari (2000)	Ref 123 (0/123) OP 83 (0/83)	Ref 66.6 (56–88) OP 70 (49–87)	Hologic QDR-4500 A	Ref: PM women from prospective population-based study OP: Women with VF Excluded scoliosis	Melton	OP: Vertebra ≥ grade 1 32 Person	OP: Vertebra ≥ grade 1 72.96 Person
Ferrari (2008)	LR 459 (459) HR 298 (0/298)	LR 68 (55–79) HR 68 (55–80)	Hologic QDR-4500 A	PM women recruited from GP Sheffield arm OPUS PM women attending MBC Sheffield Recent low-trauma fracture, VF, prednisolone >5 mg for 3 months	ABQ	LR: Vertebra ≥ grade 1 2 ≥ grade 2 NR Person ≥ grade 1 11 ≥ grade 2 NR	LR: Vertebra ≥ grade 1 70.6 85.1 ≥ grade 2 75.7 99.9 Person ≥ grade 1 71.2 97.4 ≥ grade 2 74.4 99.5
						HR: Vertebra ≥ grade 1 5 ≥ grade 2 NR Person	HR: Vertebra ≥ grade 1 79.4 99.5 ≥ grade 2 67.1 99.5 Person

Table 3 (continued)

Author, publication year	Number of patients (m/f)	Age (range)	Hardware	Inclusion criteria	Analysis method	VF prevalence (%)	Outcome parameters (sensitivity/specificity)
Fuerst (2009)	203 4(0/203)	67.5 (NR)	Hologic Delphi (n = 99) Lunar Prodigy (n = 104)	Multicenter in patients with osteoporosis	Genant	≥ grade 1 29 ≥ grade 2 20 Vertebra ≥ grade 1 4.8 ≥ grade 2 2.5 Person ≥ grade 1 26 ≥ grade 2 NR Vertebra ≥ grade 1 5 ≥ grade 2 2.3 Person ≥ grade 1 37 Vertebra ≥ grade 1 7.1 Person ≥ grade 1 32 Vertebra ≥ grade 1 11 ≥ grade 2 6.9 Person ≥ grade 1 38 ≥ grade 2 26 Vertebra	≥ grade 1 84.3 96.8 ≥ grade 2 85.7 96.4 Vertebra ≥ grade 1 70–73 99–99 ≥ grade 2 70–86 99–99 Person ≥ grade 1 72–73 93–95 ≥ grade 2 77–90 99–99 Vertebra ≥ grade 1 83.6 99.1 ≥ grade 2 94.4 99.2 Person ≥ grade 1 NR NR Vertebra ≥ grade 1 83.7 100 Person ≥ grade 1 70.6 100 Vertebra ≥ grade 1 77.4 98.4 ≥ grade 2 91.9 96.3 Person ≥ grade 1 78.3 89.0 ≥ grade 2 97.6 84.9 Vertebra
Hospers (2009)	250 (60/190)	62.0 (25–89)	Hologic Discovery	Patients referred to NMP for suspicion primary/secondary OP	Genant		
Mazzaferro (2006)	53 (31/22)	45 (23–68)	Hologic QDR-4500A	Renal transplant recipients	Eastell		
Rea (2000)	161 (0/161)	64 (49–81)	Hologic QDR-4500A	PM women recruited from GP, and PM women in MBC - T-score ≥ 2SD - 1 VF Excluded scoliosis	Visual subjective assessment		
Rud (2016)	235 (25/210)	73 (65–90)	Hologic Discovery A	Consecutive patients referred for osteoporosis assessment	Genant		
Schousboe (2006)	205 (0/205)	74.2 (65–93)	Hologic Delphi C/W	First 100: all women referred for BMD analysis ≥65 years After: women ≥65 years with osteoporosis or osteopenia	Genant	Person ≥ grade 1 44 ≥ grade 2 34 Vertebra ≥ grade 1 NR ≥ grade 2 (0.8–1.1) Person ≥ grade 1 NR ≥ grade 2 (6.9–7.8) Vertebra	Person ≥ grade 1 63 82 ≥ grade 2 73 87 Vertebra ≥ grade 1 47–57 99–99 ≥ grade 2 52–54 99–99 Person ≥ grade 1 52–63 85–89 ≥ grade 2 63–79 93–93 Vertebra
Vokes (2003)	66 (gender distribution unclear)	64 (19–93)	Lunar Prodigy	Subjects referred for routine BMD measurement	Genant	Person ≥ grade 2 32	Person ≥ grade 2 95 82

Age is expressed as mean (range)

M male, F female, VF vertebral fracture, PM postmenopausal, BMD bone mineral density, NR not reported, GP general practitioner, HIV human immunodeficiency virus, Ref reference population, OP osteoporosis, LR low risk, HR high risk, MBC metabolic bone clinic, NMP nuclear medicine department, SD standard deviation

Per vertebra analysis

Two studies did not report the VFA sensitivity and specificity per vertebra [25, 26].

The reported sensitivity of VFA to detect a vertebral fracture \geq grade 1 ranged from 46.7 to 98.7% and from 52.4 to 94.4% to detect a grade 2 or 3 vertebral fracture. The reported specificity of VFA to detect a vertebral fracture \geq grade 1 ranged from 85.1 to 99.9% and the specificity range to detect a vertebral fracture \geq grade 2 was 92 to 99.5%.

Per-person analysis

Twelve studies reported VFA parameters per patient basis [17, 19–23, 25–27, 29–31].

The VFA sensitivity range to detect a patient with a \geq grade 1 vertebral fracture was 52% to 97.2% and with a \geq grade 2 vertebral fracture was 62 to 95%. The specificity ranged from 74 to 98.9% to detect a patient with a vertebral fracture \geq grade 1 and ranged from 82 to 99% to detect a patient with a vertebral fracture \geq grade 2.

Risk of bias assessment

Seven studies were classified as having a low risk of bias [17, 19, 22, 23, 26, 29, 30]. The other nine studies were classified as having an intermediate risk of bias.

Twelve studies had no clear consecutive inclusion of patients, 4 studies did not have a per vertebra and per-person analysis of only \geq grade 2 vertebral fractures for per vertebra and per person and 3 and 6 studies did not respectively have a per-vertebra and per-person analysis. One study did not have a clear description of vertebral fractures and another study lacked clear information about blinding of observers (Supplemental Table 1).

Meta-analysis

In the meta-analysis, sensitivity and specificity were calculated per vertebra and per person (Fig. 3). In the per vertebra analysis to detect a vertebral fracture \geq grade 1, sensitivity was 0.82 (95% CI, 0.75–0.87) and specificity was 0.99 (95% CI, 0.98–1.00). In the per-person analysis to detect a vertebral fracture \geq grade 1, sensitivity was 0.85 (95% CI, 0.74–0.92) and specificity was 0.93 (95% CI, 0.87–0.97).

The per-vertebra sensitivity of VFA to detect a vertebral fracture \geq grade 2 was 0.80 (95% CI, 0.68–0.89), and specificity was 0.98 (95% CI, 0.93–0.99). The per-person sensitivity of VFA to detect patients with a vertebral fracture \geq grade 2 was 0.84 (95% CI, 0.72–0.92) and specificity was 0.90 (95% CI, 0.84–0.94).

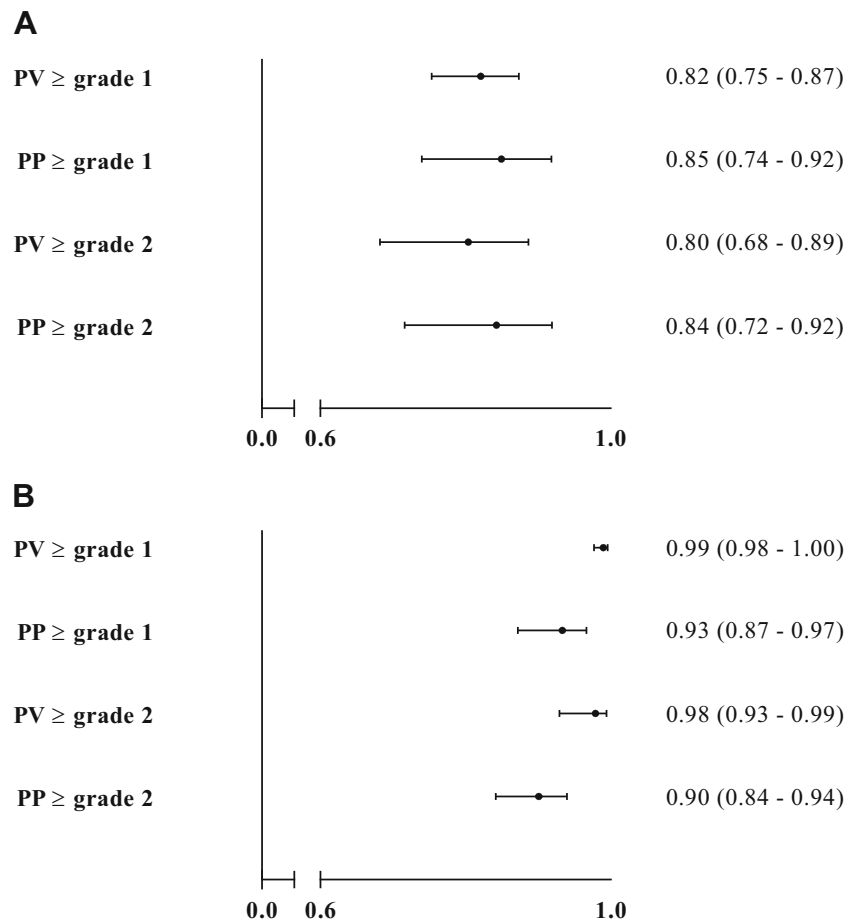
Discussion

We performed a systematic review of the literature and a meta-analysis of published data to evaluate the performance of VFA compared to conventional spine radiography in the identification of vertebral fractures in patients at high risk for osteoporosis. Findings from these data show a sensitivity of 0.82 (95% CI, 0.75–0.87) and specificity of 0.99 (95% CI, 0.98–1.00) on a per-vertebra basis and a sensitivity of 0.85 (95% CI, 0.74–0.92) and specificity of 0.93 (95% CI, 0.87–0.97) on a per-person basis. The highly variable sensitivity (47–99%) and specificity (74–100%) between reported studies is likely to be due to the wide age range, variable gender distribution and difference in recruitment of patients (from general practitioners, the outpatient clinics or from an admission ward) between studies. These differences, which were also recognized in a recent systematic review [32], represent a significant limitation in the interpretation and comparison of findings between studies.

Our meta-analysis of available data from published studies show adequate sensitivity and specificity, also when a vertebral fracture was defined as a vertebral fracture \geq grade 2: sensitivity of 0.81 (95% CI, 0.67–0.91) and specificity of 0.98 (95% CI, 0.94–1.00) for per-vertebra analysis and sensitivity of 0.84 (95% CI, 0.72–0.92) and specificity of 0.90 (95% CI, 0.84–0.94) for per-person analysis. It would be expected that performance of VFA increased if only vertebral fractures \geq grade 2 were included. However, intriguingly, the performance of VFA improved when the analysis included vertebral fractures \geq grade 1 rather than only vertebral fractures \geq grade 2. This may be explained by the fact that two of the largest published studies had excellent performance parameters and provided nearly half of all patients included in the meta-analysis of performance for identifying vertebral fractures \geq grade 1 [19, 23]. However, an analysis for the detection of vertebral fractures \geq grade 2 was not performed in these two studies, which may explain the difference in sensitivity and specificity in identifying vertebral fractures \geq grade 1 and \geq grade 2. The risk of bias assessment showed that 7 out of 16 studies had a low risk of bias, and 9 were at moderate risk of bias. It is of note, however, that the majority of these studies did not provide adequate information regarding the inclusion process of patients.

Vertebral fracture assessment has become a commonly used tool for the detection of vertebral fractures in the setting of Fracture Liaison Services, clinical care pathways where patients who have recently sustained a fracture are screened for osteoporosis and for potential underlying secondary factors for increased fracture risk. Conventional radiographs of the thoracic and lumbar spine are used as the gold standard for identification of a vertebral fracture. It has been suggested that VFA may represent an attractive alternative to spine radiographs for the detection of vertebral fractures because of the

Fig. 3 **a** Random-effects meta-analysis of sensitivity of VFA to detect vertebral fractures. *PV* per vertebra, *PP* per person. **b** Random-effects meta-analysis of specificity of VFA to detect vertebral fractures. *PV* per vertebra, *PP* per person



simplicity of the technique (using available DXA device) and lower radiation doses than those used in conventional spine radiographs. However, the advantage granted by a lower radiation dose is unfortunately counterbalanced by higher noise rates and therefore lower image quality, often precluding adequate visualization of vertebrae for the presence of a fracture. This may potentially lead to under diagnosis of vertebral fractures or the need for confirmatory spine radiography.

In our FLS, a VFA is performed in all patients at the time of BMD measurements and conventional spine radiography. We performed a retrospective study comparing low-radiation VFA with conventional spine radiography in the detection of patients with vertebral fractures \geq grade 2 in 542 men and women who had recently sustained a fracture. VFA correctly detected 77% of all patients with a vertebral fracture and correctly identified 80% as having no vertebral fracture. Low-radiation VFA was false positive in 82/410 (20%) patients who had no vertebral fractures on conventional radiography, potentially resulting in over diagnosis and thus initiating unnecessary osteoporosis treatment. Perhaps more worryingly, low-radiation VFA failed to identify a vertebral fracture \geq grade 2 or more in 30 of 132 patients (23%) and more than half of patients had ≥ 1 vertebrae that could not be evaluated by VFA, the majority of which were at the upper thoracic spine

region (level Th4 and Th5), potentially resulting in under diagnosis and under treatment. Of these, only three were missed because the fractured vertebrae were deemed unevaluable by VFA, suggesting poor diagnostic performance, the precise cause of which is as yet to be identified, rather than just poor visualization due to poor image quality.

Our study has strengths as well as limitations. Its main strength is the large group of consecutive patients of both genders all aged ≥ 50 years who had recently sustained a fracture and who were uniformly evaluated using our FLS standard protocols. A possible limitation of the study is that the inclusion of 144 patients was precluded by the lack of data on VFA or radiography. Whereas a further limitation could be the theoretical influence of a learning curve to obtain VFA images as this tool was only implemented in our FLS care pathway from 2012 onwards, we found no difference in VFA performance in the first 100 patients compared to the last 100 patients (data not shown). A matter of concern in our study is that the number of patients with ≥ 1 unevaluable vertebrae is rather high, particularly in the upper thoracic region. This problem has been reported in other VFA studies, which led Deleskog and colleagues to suggest that the method was inferior to conventional spinal radiography [18]. Notwithstanding, it appears that it may be possible to technically enhance the

performance of VFA by methods aiming at improving image quality (thus reducing the number of vertebrae that could not be visualized and improving the measurement of height loss of the vertebrae). A limiting factor in the analysis of published data is the general scarcity provided on VFA methodology, particularly radiation dosages, which may have a significant impact on the quality of obtained images. In addition, studies included in our systematic review and meta-analysis were published over a more than 15-year timeframe, spanning the years 2000 to 2016 and the improvement in hardware and software of VFA technology may have potentially influenced the outcomes. The contribution of different technologies to discrepancies in the identification of vertebral fractures has been addressed in a study comparing Lunar Prodigy and Lunar iDXA densitometers, which demonstrated that iDXA had a better performance record for visualization, and thus evaluation, of vertebrae for fractures than the Prodigy densitometer [33]. So far, there have been no studies comparing VFA performance between single-energy and dual-energy x-ray devices. The discrepancy in results of vertebral fracture assessment using VFA compared to conventional radiology in our study is in contrast to the concordance of results of assessments between radiology and VFA in the majority of studies reported in the systematic review and used in the meta-analysis. This difference could have been influenced by the different methodology used between studies. Quantitative assessment was thus used to evaluate VFA images in our study compared to the use of Genant's semi-quantitative assessment in the vast majority of studies included in the systematic review and meta-analysis.

In conclusion, from our meta-analysis, findings of published data demonstrate adequate performance parameters of VFA in studies designed for patients at risk for osteoporosis, although a limitation was the very broad range of prevalent vertebral fractures (6.9–100%) and age (23–96 years) which may have influenced study outcomes. The data of our FLS study were in contrast with the numbers of the meta-analysis. The precise cause of the underperformance of VFA in our center is currently being investigated. Our findings suggest that caution should be advocated with the interpretation of VFA data and that centers should check the performance of their VFA device against conventional radiography of the spine before exclusively relying on this tool in the identification of vertebral fractures.

Acknowledgements Thanks to Petra Dibbets-Schneider for her help with the analyses of VFA images and to Johannes W. Schoones for his help with the search strategy for the systematic review.

Compliance with ethical standards

Conflicts of interest None.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Center JR, Nguyen TV, Schneider D, Sambrook PN, Eisman JA (1999) Mortality after all major types of osteoporotic fracture in men and women: an observational study. *Lancet* 353:878–882
- Romagnoli E, Carnevale V, Nofroni I, D'Erasmo E, Paglia F, De Geronimo S, Pepe J, Raejntroph N, Marangi M, Minisola S (2004) Quality of life in ambulatory postmenopausal women: the impact of reduced bone mineral density and subclinical vertebral fractures. *Osteoporos Int* 15:975–980
- Cummings SR, Melton LJ (2002) Epidemiology and outcomes of osteoporotic fractures. *Lancet* 359:1761–1767
- Johnell O, Kanis JA (2006) An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. *Osteoporos Int* 17:1726–1733
- Black DM, Arden NK, Palermo L, Pearson J, Cummings SR (1999) Prevalent vertebral deformities predict hip fractures and new vertebral deformities but not wrist fractures. Study of Osteoporotic Fractures Research Group. *J Bone Miner Res* 14:821–828
- Roux C, Fechtenbaum J, Kolta S, Briot K, Girard M (2007) Mild prevalent and incident vertebral fractures are risk factors for new fractures. *Osteoporos Int* 18:1617–1624
- Delmas PD, Genant HK, Crans GG, Stock JL, Wong M, Siris E, Adachi JD (2003) Severity of prevalent vertebral fractures and the risk of subsequent vertebral and nonvertebral fractures: results from the MORE trial. *Bone* 33:522–532
- Delmas PD, van de Langerijt L, Watts NB, Eastell R, Genant H, Grauer A, Cahall DL (2005) Underdiagnosis of vertebral fractures is a worldwide problem: the IMPACT study. *J Bone Miner Res* 20: 557–563
- Cooper C, Melton LJ (1992) Vertebral fractures. *BMJ* 304:793–794
- Genant HK, Wu CY, van Kuijk C, Nevitt MC (1993) Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res* 8:1137–1148
- Genant HK, Li J, Wu CY, Shepherd JA (2000) Vertebral fractures in osteoporosis: a new method for clinical assessment. *J Clin Densitom* 3:281–290
- CBO (2011) Richtlijn Osteoporose en Fractuurpreventie. <http://www.diliguide.nl/document/1015> (accessed 12–6–2015)
- Kanis JA, McCloskey EV, Johansson H, Cooper C, Rizzoli R, Reginster JY (2013) European guidance for the diagnosis and management of osteoporosis in postmenopausal women. *Osteoporos Int* 24:23–57
- Malgo F, Appelman-Dijkstra NM, Termaat MF, van der Heide HJ, Schipper IB, Rabelink TJ, Hamdy NAT (2016) High prevalence of secondary factors for bone fragility in patients with a recent fracture independently of BMD. *Arch Osteoporos* 11:12
- Leefflang MM (2014) Systematic reviews and meta-analysis of diagnostic test accuracy. *Clin Microbiol Infect* 20:105–113
- Ferrar L, Jiang G, Barrington NA, Eastell R (2000) Identification of vertebral deformities in women: comparison of radiological assessment and quantitative morphometry using morphometric radiography and morphometric X-ray absorptiometry. *J Bone Min Res* 15: 575–585

17. Ferrar L, Jiang G, Clowes JA, Peel NF, Eastell R (2008) Comparison of densitometric and radiographic vertebral fracture assessment using the algorithm-based qualitative (ABQ) method in postmenopausal women at low and high risk of fracture. *J Bone Min Res* 23:103–111
18. Deleskog L, Laursen NO, Nielsen BR, Schwarz P (2016) Vertebral fracture assessment by DXA is inferior to X-ray in clinical severe osteoporosis. *Osteoporos Int* 27:2317–2326
19. Diacinti D, Guglielmi G, Pisani D, Diacinti D, Argiro R, Serafini C, Romagnoli E, Minisola S, Catalano C, David V (2012) Vertebral morphometry by dual-energy X-ray absorptiometry (DXA) for osteoporotic vertebral fractures assessment (VFA). *Radiol Med* 117:1374–1385
20. Mazzaferro S, Diacinti D, Proietti E, Barresi G, Baldinelli M, Pisani D, D'Erasmus E, Pugliese F (2006) Morphometric X-ray absorptiometry in the assessment of vertebral fractures in renal transplant patients. *Nephrol Dial Transplant* 21:466–471
21. Damiano J, Kolta S, Porcher R, Tournoux C, Dougados M, Roux C (2006) Diagnosis of vertebral fractures by vertebral fracture assessment. *J Clin Densitom* 9:66–71
22. Bazzocchi A, Spinnato P, Fuzzi F, Diano D, Morselli-Labate AM, Sassi C et al (2012) Vertebral fracture assessment by new dual-energy X-ray absorptiometry. *Bone* 50:836–841
23. Diacinti D, Del Fiacco R, Pisani D, Todde F, Cattaruzza MS, Diacinti D, Arima S, Romagnoli E, Pepe J, Cipriani C, Minisola S (2012) Diagnostic performance of vertebral fracture assessment by the lunar iDXA scanner compared to conventional radiography. *Calcif Tissue Int* 91:335–342
24. Hospers IC, van der Laan JG, Zeebregts CJ, Nieboer P, Wolffenbuttel BH, Dierckx RA, Kreeftenberg HG, Jager PL, Slart RH (2009) Vertebral fracture assessment in supine position: comparison by using conventional semiquantitative radiography and visual radiography. *Radiology* 251:822–828
25. Vokes TJ, Dixon LB, Favus MJ (2003) Clinical utility of dual-energy vertebral assessment (DVA). *Osteoporos Int* 14:871–878
26. Rud B, Vestergaard A, Hyldstrup L (2016) Accuracy of densitometric vertebral fracture assessment when performed by DXA technicians—a cross-sectional, multiobserver study. *Osteoporos Int* 27:1451–1458
27. Chapurlat RD, Duboeuf F, Marion-Audibert HO, Kalpakcioglu B, Mitlak BH, Delmas PD (2006) Effectiveness of instant vertebral assessment to detect prevalent vertebral fracture. *Osteoporos Int* 17:1189–1195
28. 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
29. Fuerst T, Wu C, Genant HK, von Ingersleben G, Chen Y, Johnston C, Econs MJ, Binkley N, Vokes TJ, Crans G, Mitlak BH (2009) Evaluation of vertebral fracture assessment by dual X-ray absorptiometry in a multicenter setting. *Osteoporos Int* 20:1199–1205
30. 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
31. Rea JA, Li J, Blake GM, Steiger P, Genant HK, 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
32. Lee JH, Lee YK, Oh SH, Ahn J, Lee YE, Pyo JH, Choi YY, Kim D, Bae SC, Sung YK, Kim DY (2016) A systematic review of diagnostic accuracy of vertebral fracture assessment (VFA) in postmenopausal women and elderly men. *Osteoporos Int* 27:1691–1699
33. 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