

Diagnosis and grading of radiographic osteoporotic vertebral deformity by general radiologists after a brief self-learning period

Yì Xiáng J. Wáng¹, Wei-Hong Liu², Davide Diacinti^{3,4}, Da-Wei Yang⁵, Antonio Iannaccone⁶, Xiao-Rong Wang⁷, Endi Kripa⁶, Nazmi Che-Nordin¹, Daniele Diacinti⁶

¹Department of Imaging and Interventional Radiology, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, China; ²Department of Radiology, General Hospital of China Resources & Wuhan Iron and Steel Corporation, Wuhan, China; ³Department of Oral and Maxillofacial Sciences, Umberto I Hospital, Sapienza University of Rome, Rome, Italy; ⁴Department of Diagnostic and Molecular Imaging, Radiology and Radiotherapy, University Foundation Hospital Tor Vergata, Rome, Italy; ⁵Department of Radiology, Beijing Friendship Hospital, Capital Medical University, Beijing, China; ⁶Department of Radiological Sciences, Oncology and Pathology, Sapienza University of Rome, Rome, Italy; ⁷Department of Radiology, Ningbo First Hospital, Ningbo, China

Contributions: (I) Conception and design: YXJ Wáng; (II) Administrative support: NC Che-Nordin, D Diacinti; (III) Provision of study materials or patients: YXJ Wáng, D Diacinti; (IV) Collection and assembly of data: YXJ Wáng, NC Che-Nordin, D Diacinti; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Dr. Yì Xiáng J. Wáng. Department of Imaging and Interventional Radiology, the Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, China. Email: yixiang_wang@cuhk.edu.hk.

Background: The expanded semi-quantitative (eSQ) osteoporotic vertebral deformity (OVD) classification has minimal, mild, moderate, moderately-severe, severe, and collapsed grades with <20%, 20–25%, >25%–1/3, >1/3–40%, >40%–2/3, >2/3 vertebral height loss respectively. This study evaluates the performance of using this grading criterion by radiology readers who did not have former training in OVD assessment.

Methods: Spine radiographs of 44 elderly women with 278 normal appearing vertebrae and 65 OVDs were selected, with two senior readers agreed the reference reading. Three readers from Italy and three readers from China were invited to evaluate these radiographs after reading five reference articles including one detailing eSQ criteria with illustrative examples. Before the second round of reading, the readers were asked to read an additional explanatory document. For the readers in Italy an additional on-line demonstration was given on how to measure vertebral height loss in another five cases of OVD. Two Chinese readers had a third round of reading after a 90 minutes' on-line lecture.

Results: The final absolute agreement rate with the reference reading (i.e., exactly the same grading as the reference) ranged between 46.2% to 68.2% for the six readers, and the final relative agreement (with one eSQ grade difference allowed) ranged between 78.5% to 92.5%. The >1 grade disagreement rate was all below 11%, and mostly below 7%. The missed OVD were mostly minimal grade. The rate for missing a ≥ mild OVD was <4.5%, and false positive rate was generally <1.4% among the final reading. If the minimal grade was removed and the remaining gradings were converted to Genant's semi-quantitative (GSQ) grading, the mean kappa values against the reference reading for SQ grades-1,2,3 were 0.813, 0.814, and 0.916 respectively.

Conclusions: This study demonstrates good performance of the six learner readers for assessing radiographic after a brief self-learning period.

Keywords: Osteoporosis; osteoporotic vertebral deformity (OVD); osteoporotic vertebral fracture (OVF); spine; grading

Submitted Jul 07, 2020. Accepted for publication Aug 06, 2020.

doi: 10.21037/jtd-20-2379

View this article at: <http://dx.doi.org/10.21037/jtd-20-2379>

Introduction

Osteoporotic vertebral fracture (OVF) is the most common osteoporotic fracture. A vertebral fracture, after minor trauma, is a hallmark of osteoporosis. OVF predispose to both further fracturing and fracture cascade (1-5). OVF can be associated with decreases in trunk extension torque, spinal motion, functional reach, mobility skills and walking distance. OVF may also influence mortality because of its association with chronic back pain, immobility and change in posture, which may increase the risk of infection and death (2). Appropriate interventions for OVF can reduce the occurrence of hip fractures, preventing further OVF and reducing pain and related disabilities. It is important to identify and report OVF, so that appropriate investigation and treatment can be instigated. However, OVF may often be relatively asymptomatic so that, in the absence of localizing symptoms, spinal fracture status is usually determined by obtaining lateral radiographs of the spine either by radiograph or by dual-energy X-ray absorptiometry (DXA) devices operating in imaging mode (4).

The Genant's semi-quantitative (GSQ) criteria is currently the most used approach for identifying OVF (5-7). The qualitative diagnosis with radiological knowledge helps to minimize false positive readings due to vertebral physiological wedging, degenerative wedging, and other congenital anomalies. Radiological evaluation can also detect subtle changes that may be clinically important, such as endplate and/or cortex fracture (ECF) (8-13). According to Genant *et al.*, a vertebra body (VB) is graded on visual inspection of the anterior, middle, or posterior heights as normal (grade-0), mildly deformed (grade-1, a 20–25% reduction in one of the three heights and a reduction in area of 10–20%), moderately deformed (grade-2, a 25–40% reduction in any height and a reduction in area of 20–40%) and severely deformed (grade-3, a 40% or more reduction in height and area). In addition to the height reduction estimation, emphasis is also placed on the radiological evaluation of osteoporotic vertebral deformity (OVD). However, unless there is a face-to-face training with experienced readers, it has been well documented that it is difficult to apply GSQ criteria by only reading the text description of Genant *et al.* (5,14-17). For example, despite that a standardized protocol of radiograph acquisition techniques and of interpretation criteria was applied, Diacinti *et al.* (16) reported a study that, among 562 OVF identified by radiologist readers in peripheral hospitals, 102 were classified as normal vertebrae by the experienced

radiologist readers in a central hospital; while 205 OVF were incorrectly evaluated by local readings as (false) negatives.

Built upon the GSQ criteria, Wang *et al.* (18) recently proposed an expanded semi-quantitative (eSQ) OVD classification with the following features: (I) minimal grade is OVDs with <20% VB height loss; (II) eSQ mild grade is the same as GSQ mild grade (20–25% height loss); (III) GSQ moderate grade is subdivided into eSQ moderate grade (mod/m, >25%–1/3 height loss) and eSQ moderately-severe grade (mod/s, >1/3–40% height loss); (IV) GSQ severe grade is subdivided into eSQ severe grade (>40%–2/3 height loss) and eSQ collapsed grade (with >2/3 height loss). The eSQ criteria introduce a formal minimal grade, equivalent to GSQ grade-0.5, referring to radiological OVDs with less than 20% height loss. While GSQ grade-0.5 OVDs are common, it is less reported and discussed in literature (6). eSQ minimal grade can be clinically relevant specially when there is associated ECF (8-12,19). GSQ moderate is subdivided into eSQ mod/m and mod/s grades because OVDs with >1/3 height loss are always associated with positive ECF sign radiographically (this is the reason for such a subdivision) (12,20,21). It has been noted that compared with VBs without OVD, ECF(-) GSQ grade-1 and grade-2 OVDs do not have a higher short-term future risk for new incident OVD. Within the same GSQ mild/moderate OVD grades, compared with the subjects without ECF, the subjects with ECF are associated with a higher short-term future risk of OVD progression and new incident OVD (12). A subdivision GSQ grade-3 will allow recording of severe OVD progressions (such as a 45% height loss progresses to 75% height loss). Wang *et al.* (18) defined how VB height loss is measured, and also provide graphic illustration of 36 cases with radiographic OVD.

This study aims to test the hypothesis that, since Wang *et al.*'s article defines how VB height loss is measured and users can refer to the 36 cases illustrated in the article, a general radiologist can self-learn by reading a couple of key reference papers including Wang *et al.*'s, and performs well in evaluating radiographic OVD without face-to-face coaching from experienced readers.

Methods

This study was approved by the local ethics committees of authors' institutions. Informed consent for additional analysis of radiographs in this study was waived. The

Table 1 Vertebral height loss criteria for Genant semi-quantitative (GSQ) grading and extended semi-quantitative grading (eSQ)*

Variable	Extent of vertebral height loss					
	<20%	≥20–25%	>25%–1/3	>1/3–40%	>40%–2/3	>2/3
Genant SQ	Grade-0.5	Grade-1	Grade-2		Grade-3	
Genant SQ		Mild	Moderate		Severe	
Extended SQ	Grade-1	Grade-2	Grade-3	Grade-4	Grade-5	Grade-6
Extended SQ	Minimal (min)	Mild (mil)	Moderate (mod/m)	Moderately-severe (mod/s)	Severe (sev)	Collapsed (col)

* , vertebrae with normal radiographical morphology is noted as grade-0.

thoracic and/or lumbar spine radiograph of 29 Chinese cases and 15 Italian cases were randomly selected from senior authors' study database, with in total 343 VBs visible on the radiographs. All subjects were females and aged >55 years old, and according to the two senior authors (YXJW, DD) had at least one OVD per subject according to the eSQ criteria. The two senior authors had been involved in the OVD/OVF evaluation in the past, but they did not have face-to-face interaction. These two senior authors initially evaluated the 44 cases independently according to eSQ criteria (Table 1), and found they agreed in 278 VBs without OVD, and 65 VBs with OVD among them there were 7 OVDs with disagreement by one eSQ grade (none had disagreement by > one grade), and one VB with upper endplate Schmorl's node co-existing with ambiguous minimal grade OVD. A second-round reading resolved disagreement in 4 OVDs, and with remaining three OVDs considered two grades were acceptable for each VB (e.g., an OVD with a measured height loss of 39% was considered acceptable for both mod/s and severe grades). The agreed reading of these two readers were then regarded as the reference. In total, there were 10 minimal, 11 mild, 21 mod/m, 7 mod/s, 17 severe, and 3 collapsed grade OVDs; while one OVD was either minimal grade or mild grades, two OVDs were either mod/s grade or severe grades; and for additional one VB, both minimal grade OVD or without OVD were considered acceptable.

This study recruited three young general radiologist readers from Italy (I1, I2, I3) and three young general radiologists from China (C1, C2, C3). They were asked not to communicate on the study topics during the testing period. None of them were subspecialized in musculoskeletal radiology and none of them participated special clinical program in OVD assessment. These 6 radiologists were asked to read the three papers of references 5, 18, 22 as the primary teaching materials (5,18,22), and

two papers of references 23, 24 as supplementary teaching materials (23,24). Among them Wang *et al.*'s article is the key teaching material (18). Szulc's letter gives a good introduction of how to identify an OVD (5), and Wang & Che-Nordin's letter explains Genant's SQ criteria (22). The two supplementary papers mainly provide teaching material on ECF identification, and in the meantime provide more examples of OVD (23,24). After reading the teaching articles, the 6 readers were asked to evaluate the 44 study cases' blinded radiographs, identifying the VBs with OVD and their eSQ grading. The VB dimension measurement was done using the software MicroDicom DICOM viewer (<https://www.microdicom.com/>). The measurement was computer screen-captured and feedback to the two senior authors. For second round of reading, the readers in China were asked to read an additional document (supplementary document at: <http://fp.amegroups.cn/cms/fae5c5f22eef21ad80f0b8003d120288/JTD-20-2379-1.pdf>), and then to evaluate the radiographs again. This additional document was prepared after senior author-YXJW reviewed the reading results of the 6 readers with feedback of computer screen-captures. This document comments on the common errors of the first round of test readings, but no specific case or example was discussed. For the readers in Italy, in addition to reading this document, senior author-DD did an on-line demonstration of how to measure vertebral height loss in another 5 five cases of OVD which the readers in Italy measured two days later, and then experienced senior author-DD corrected their measurements and gave feedback to the three Italian readers. After that, the Italian readers evaluated the radiographs again. Two Chinese readers (C2, C3) had a third round of reading. Via an on-line platform, senior author-YXJW spent 90 minutes to explain Wang *et al.*'s article (18); and explained his approach to evaluate and grade OVD. The content of this lecture focuses on the important points of what were already on

Table 2 Final agreement (vertebral number and percentage) between the six learner readers and reference reading

Variable	OVD absolute agree*	OVD relative agree*	OVD disagree by >1 grade*	≥ mild OVD missed*	False (+) OVD [#]
I1	43 (66.2%)	57 (87.7%)	2 (3.1%)	1/1.6%	1/0.36%
I2	39 (60%)	54 (83.1%)	4 (6.2%)	0	0
I3	30 (46.2%)	51 (78.5%)	7 (10.8%)	0	0
C1	42 (63.6%)	60 (90.9%)	4 (6.1%)	0	4 (1.4%)
C2	45 (68.2%)	61 (92.5%)	0	3 (4.5%)	0
C3	42 (63.6%)	59 (89.4%)	4 (6.1%)	2 (3.0%)	2 (0.7%)

OVD absolute agree: exactly the same grading as the reference reading; OVD relative agree: compared with the reference reading, one grade difference is allowed; OVD disagree by >1 grade: disagreement by more than one grade, for example, a mild grade is marked as mod/s grade; ≥ mild OVD missed: mild grade OVD were missed with C2 missed one severe OVD (due to oversight). *, out of total true OVD; [#], falsely marked as OVD out of total vertebrae with no OVD. OVD, osteoporotic vertebral deformity.

the teaching materials, with no additional case or example provided.

For statistical analysis, Kappa coefficient was used for to measure inter-reader reliability for categorical items. The performance of the 6 general radiologists was assessed with agreement criteria comparing with the reference reading by the two senior authors.

Results

The key indicators of final agreement between the six learner readers and reference reading are summarized in *Table 2*, and the detailed agreements are shown in *Table 3*. *Table 2* shows the final absolute agreement rate (i.e., exactly the same grading as the reference) ranged between 46.2% to 68.2%, and the final relative agreement (with one grade difference allowed) ranged between 78.5% to 92.5%. The >1 grade disagreement in grading was all below 11%. The missed OVDs were mostly minimal grade. The rate for missing a ≥ mild OVD was <4.5%. One severe OVD was missed by Chinese reader-2; however, it was an oversight rather than misjudgement which was confirmed after the study completed. The false positive rate was generally <1.4% among the final reading, and all were marked as minimal grade. If grading was disregarded and a VB was only considered either OVD positive or negative, according to eSQ the positive agreement was 90.8%, 89.2%, 89.2%, 96.9%, 92.4%, and 96.9% respectively for the six learner readers; and for negative VBs, the agreement was 99.6%, 100%, 100%, 98.2%, 98.9%, and 98.9% respectively.

If the eSQ minimal grade was removed, eSQ mod/m and mod/s grades were merged into a single moderate grade, and eSQ severe and collapsed grades were merged into

severe grade, the reading results were then converted into GSQ results. The GSQ results are shown in *Table 4*. The mean kappa value for GSQ grades-1,2,3 was 0.813, 0.814, and 0.916, respectively.

For the second round of reading, the three Italian readers and Chinese reader-1 performed the same as the 1st round of reading, while Chinese reader-2 and -3 improved notably. Chinese reader-2 further improved during the 3rd round of reading.

Discussion

Despite years' of research, the radiographic criteria for diagnosing osteoporotic VF and its grading remain debated (14,25-28). Recent evidences favor radiological diagnosis over OVD's morphometric criteria (14,27,28). For example, Leidig-Bruckner *et al.* (27) applied an algorithm for radiological differential classification (RDC), and their RDC's reading were compared with BMD measurement as well as Eastell and McCloskey's morphometric criteria (29,30). In women, 31-68% of cases with morphometric OVD were classified by RDC as non-osteoporotic. On the other hand, up to 48% OVD in women were missed by morphometry. Although OVD by most definitions were significantly and inversely related to BMD, a stronger association existed between BMD and osteoporotic fractures defined by RDC (17). In the early 1990s, Genant *et al.* (6) proposed the GSQ criteria. Though the GSQ is commonly used, it is associated with two major difficulties. The first is to decide whether a mild OVD exists for a VB, or it is only a degenerative wedging. The second is the precise grading. These difficulties may not be so an issue for clinical practice; but can cause problems for epidemiological

Table 3 Detail agreement (vertebral number) between the six learner readers and reference reading

Readers	Negative VB	Agreed OVD	OVD disagreed by 1 grade	Agreed OVD including 1 grade disagreement	OVD disagree by >1 grade	Minimal OVD missed	Mild OVD missed	Severe OVD missed	False (+) OVD marked minimal
I1-1	277	43	14	57	2	5	1	0	1
I1-2	277	43	14	57	2	5	1	0	1
I2-1	278	39	15	54	4	7	0	0	0
I2-2	278	39	15	54	4	7	0	0	0
I3-1	278	30	21	51	7	7	0	0	0
I3-2	278	30	21	51	7	7	0	0	0
C1-1	276	42	18	60	4	1	0	0	2
C1-2	274	42	18	60	4	1	0	0	4
C2-1	270	41	20	61	0	2	2	1	7
C2-2	275	44	17	61	0	2	2	1	2
C2-3	277	45	16	61	0	2	2	1	0
C3-1	277	35	18	53	3	6	3	0	1
C3-2	275	44	15	59	4	1	2	0	2
C3-3	275	42	17	59	4	1	2	0	2

VB, vertebral body; OVD, osteoporotic vertebral deformity; C2-1, the 1st round of reading by reader C2; C2-2, the 2nd round of reading by reader C2; C2-3, the 3rd round of reading by reader C2. others the same.

Table 4 Agreement of the six learner readers and reference reading when the eSQ readings were converted into Genant SQ readings*

Readers	Negative VB	Mild OVD			Moderate OVD			Severe OVD								
		Agreed	Overgraded	Missed	Agreed	Undergraded	Overgraded	Agreed	Undergraded	Missed	Kappa					
I1	288	9	1	1	18	6	2	0.897	18	6	2	0.805	18	0	0	1
I2	289	9	1	0	18	4	3	0.946	18	4	3	0.825	17	2	0	0.941
I3	288	5	5	0	16	3	7	0.659	16	3	7	0.745	18	1	0	0.971
C1	287	10	2	0	21	4	0	0.906	21	4	0	0.906	16	3	0	0.909
C2	286	6	3	2	22	6	0	0.697	22	6	0	0.870	12	5	1	0.790
C3	287	7	2	2	15	10	0	0.771	15	10	0	0.734	16	4	0	0.882
Mean								0.813				0.814				0.916

*, eSQ minimal OVD is generally removed for the conversion. However, in a few cases when a mild grade was marked as minimal grade, it is considered to be 'agreed' rather than missed; and in a few cases when a minimal grade was marked as mild grade, it is considered to be 'agreed' rather than false positive. Due to the existence of four VB had ambiguous grading by the referee reader, OVD numbers do not exactly match among the 6 learner readers. VB, vertebral body; OVD, osteoporotic vertebral deformity.

studies, and particularly for cross-center results comparison. GSQ has been generally perceived difficult to implement by those readers who have not received training from experienced readers. The agreement among readers for OVD GSQ grading can be poor (14). GSQ method relies on estimate of vertebral dimensions, which is a potential source of observer error. Though it has not been specifically noted, it is implied that the posterior VB height is used as the reference to estimate middle/anterior height loss. However, to compare the height loss of a vertebra with its adjacent vertebrae is more reasonable (18), as the ratio of anterior height to posterior height varies among different vertebral levels. Additionally, due to the existence of vertebra's posterior lip (uncinate process), the posterior height measure can cause inconsistencies.

Wang *et al.*'s work tries to address the difficulties of GSQ. In addition to proposing the eSQ scheme with the perceived advantages, Wang *et al.*'s paper also defines the way the VB height loss is measured. It may help to resolve difference by individual readers to facilitate consensus reading. For example, a 38% height loss can be read as GSQ moderate grade by one reader and GSQ severe grade by another reader, which would be difficult to reconcile without a consistent measurement. Since original GSQ criteria were primarily explained by text (6,7), rather than by examples, Wang *et al.*'s paper provides 36 illustrative examples with eSQ and GSQ grading. Though Genant *et al.* described GSQ grade-0.5 (a qualitative OVD with less than 20% height loss) (6), it was rarely used and reported, and to our knowledge many GSQ grade-0.5 OVDs might have been assigned as SQ grade-1 in literature by some readers (and may be ignored by other readers) (6,31). The introduction of a formal eSQ grade-1 may solve the problem when a qualitative OVD exists but it does not look like achieving the threshold of $\geq 20\%$ vertebral height loss. Thus, the reader is not forced to make a 'YES' or 'NO' choice. It has been noted that even some of the GSQ mild grade VD may not be true OVF (32,33). The eSQ criteria will allow flexibility in application, for example, a researcher may decide to exclude all eSQ minimal grade OVDs, or may only account eSQ minimal grade OVDs with positive ECF sign (18).

We consider the results of this study being satisfactory. With some degree of commitment, the initial reading results can be improved as shown by the 1st and 2nd round results of readers C2 and C3. A precise comparison with previous results would be difficult. Firstly, the readers in this study would have guessed that at least an OVD exists

per subject (though they were not informed as such). In real practice, readers have to read two radiographs (one thoracic spine radiograph and one lumbar spine radiograph) per subject, while, with this study, focused radiographs with on average 8 VBs each were provided. However, results in this study compare favorably with previous publications. Recently, Diacinti *et al.* (16) reported a study to evaluate the diagnostic accuracy of OVF by radiologist readers in peripheral hospitals with ordinary experience and using the GSQ criteria, their readings were compared with the reference reading by two experienced central radiologists. It was noted that 30.8% of GSQ grade-1 OVFs were mis-diagnosed by the general radiologists, despite that a standardized protocol of radiograph acquisition techniques and interpretation criteria was applied. In that study, among 562 VFs identified by general radiologist readers, 102 were classified as normal vertebrae by the central readers; while 205 vertebrae were incorrectly evaluated by general radiologist as not presenting OVFs. Fechtenbaum *et al.* (15) reported the agreement for OVF diagnosis among the 390 vertebrae in male patients between thirteen experimented rheumatologists. After the group discussed in a workshop and established an algorithm of VF diagnosis, the number of OVF detected varies from 18 to 59. The mean concordance percentage (negative or positive agreement) between experts at the vertebral level ranged from 83.1% to 97.9%. Wu *et al.* (34) reported three experienced readers' inter-reader kappa scores based on the dichotomous fracture/non-fracture decision, which ranged from 0.84 to 0.87 between the individual readings and the consensus reading, and 0.69 agreement between two readers. Buckens *et al.* (35) assessed the agreement of four observers with different levels of experience, and sagittal reformatted CT images were used. For fracture presence the interobserver kappa scores ranged from 0.56 to 0.81. For GSQ grade-3 OVF, the interobserver kappa scores ranged from 0.73 to 0.88.

One aspect of the results in this study which surprised the senior authors is that the on-line coaching did not substantially improved the reading results. This may suggest that a 1–2 hours on-line interaction is insufficient to systematically improve readers' performance. Note that despite senior authors YXJW and DD did not have face-to-face interaction, they achieved a very good agreement which appears to be even better than the results of the inter-reader agreements reported by Wu *et al.* (34) from the University of California San Francisco (USA) group who conceived the original GSQ criteria. We believe this re-emphasizes the importance of well-defined measurement method. Our

results also suggest that for the 6 young general radiologists to achieve the performance of the senior authors, a much longer period of coaching may be required. It has been noted that eSQ minimal and mild grade OVDs are often under-recognized or missed in clinical practice (36), while after self-learning the number of missed minimal and mild grade OVDs by the six learner reader were very few in this study. Thus, in addition to research purpose, Wang *et al.*'s article (18) discussed in this study may also service as a teaching material for clinical radiologists reading spine radiographs. Osteoporosis is a growing problem in the aging population, it is estimated that up to one in three women and one in five men over 50 years of age are affected (37), leading to millions of fractures annually and contributing substantially to morbidity and mortality. Subclinical OVF is an early sign of osseous fragility (4).

There are a few limitations for this study. A ground truth for OVD diagnosis and grading does not exist. Thus, the reference reading could still have suffered from subjectivity. To avoid potential controversies, readers were not asked to specifically annotate ECF sign, though we did ask readers to grade normal height VB with apparent ECF signs as minimal grade. We consider the issue of inter-reader agreement of ECF is beyond the scope of the current study, but it does have an implication in our choosing the phrasing of OVD rather OVF. In assessing a vertebra, there are usually three stepwise questions: (I) does a VD exist? (II) is this VD likely an OVD? (III) can this OVD be called OVF? We chose to use OVD in this paper as some of the minimal OVDs may not have clinical sequence (32,33). Some authors suggested that OVD is OVF only when diagnosed on the basis of evidence or findings of endplate, cortical, or trabecular damage (13,38). Imaging research is constrained by the investigators' vigilance. There is a possibility that if the participants of this study felt they were the PI (principle investigator) and fully responsible for the study (rather than being a participant), their performance might have been even better. Another point of note is that reader C2 had performed a study on traumatic VF and thus might have better knowledge of VF (39). However, during the 1st round of testing reading, reader C2 did not perform notably better than reader C1 and C3, thus unlikely this would have majorly affected the results.

In conclusion, this study tested the performance of evaluating OVD in 44 elderly women's spine radiograph by young general radiologists who did not have former experience in systematically evaluating OVD, and demonstrated good performance by these six learner

readers after reading a few key references. Missed OVD were mostly minimal grade, missed mild grade OVD was very few; and false positive rate was also low. Our results are remarkably good considering that it is generally perceived that, without training with experienced readers, SQ criteria is difficult to apply. Our results confirm that it is important to publish standardized reference materials so that radiology readers can read and apply the same OVD criteria and derive consistent readings.

Acknowledgments

Funding: This study is partially supported by ITF project of Hong Kong SAR (code: ITS/334/18).

Footnote

Data Sharing Statement: Available at <http://dx.doi.org/10.21037/jtd-20-2379>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/jtd-20-2379>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was approved by the local ethics committees of authors' institutions. Informed consent for additional analysis of radiographs in this study was waived.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Szulc P, Bouxsein ML. Overview of osteoporosis: epidemiology and clinical management. Vertebral Fracture Initiative Resource Document. Nyon,

- Switzerland: International Osteoporosis Foundation; 2011. PART I. p. 1-65.
2. Ismail AA, O'Neill TW, Cooper C, et al. Mortality associated with vertebral deformity in men and women: results from the European Prospective Osteoporosis Study (EPOS). *Osteoporos Int* 1998;8:291-7.
 3. Broy SB. The vertebral fracture cascade: etiology and clinical implications. *J Clin Densitom* 2016;19:29-34.
 4. Schwartz EN, Steinberg D. Detection of vertebral fractures. *Curr Osteoporos Rep* 2005;3:126-35.
 5. Szulc P. Vertebral fracture: diagnostic difficulties of a major medical problem. *J Bone Miner Res* 2018;33:553-9.
 6. Genant HK, Wu CY, van Kuijk C, et al. Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res* 1993;8:1137-48.
 7. Genant HK, Jergas M. Assessment of prevalent and incident vertebral fractures in osteoporosis research. *Osteoporos Int* 2003;14:S43-55.
 8. Sugita M, Watanabe N, Mikami Y, et al. Classification of vertebral compression fractures in the osteoporotic spine. *J Spinal Disord Tech* 2005;18:376-81.
 9. Jiang G, Eastell R, Barrington NA, et al. Comparison of methods for the visual identification of prevalent vertebral fracture in osteoporosis. *Osteoporos Int* 2004;15:887-96.
 10. Wáng YXJ, Che-Nordin N. Some radiographically 'occult' osteoporotic vertebral fractures can be evidential if we look carefully. *Quant Imaging Med Surg* 2019;9:1992-5.
 11. Lentle BC, Berger C, Probyn L, et al. Comparative analysis of the radiology of osteoporotic vertebral fractures in women and men: cross-sectional and longitudinal observations from the Canadian Multicentre Osteoporosis Study (CaMos). *J Bone Miner Res* 2018;33:569-79.
 12. Wáng YX, Che-Nordin N, Deng M, et al. Osteoporotic vertebral deformity with endplate/cortex fracture is associated with higher further vertebral fracture risk: the Ms. OS (Hong Kong) study results. *Osteoporos Int* 2019;30:897-905.
 13. Lentle B, Trollip J, Lian K. The Radiology of Osteoporotic Vertebral Fractures Redux. *J Clin Densitom* 2016;19:40-7.
 14. Lentle B, Koromani F, Brown JP, et al. The Radiology of Osteoporotic Vertebral Fractures Revisited. *J Bone Miner Res* 2019;34:409-18.
 15. Fechtenbaum J, Briot K, Paternotte S, et al. Bone section of the French Rheumatology Society. Difficulties in the diagnosis of vertebral fracture in men: agreement between doctors. *Joint Bone Spine* 2014;81:169-74.
 16. Diacinti D, Vitali C, Gussoni G, et al. Misdiagnosis of vertebral fractures on local radiographic readings of the multicentre POINT (prevalence of osteoporosis in INternal medicine) study. *Bone* 2017;101:230-5.
 17. Messina C, Albano D, Rapisarda S, et al. Misdiagnosis of vertebral fractures on plain films: Are radiologists really working so bad? *Bone* 2017;105:307.
 18. Wáng YX, Diacinti D, Yu W, et al. Semi-quantitative grading and extended semi-quantitative grading for osteoporotic vertebral deformity: a radiographic image database for education and calibration. *Ann Transl Med* 2020;8:398.
 19. Yoshida T, Nanba H, Mimatsu K, et al. Treatment of osteoporotic spinal compression fractures. Conservative therapy and its limitation. *Clin Calcium* 2000;10:53-8.
 20. Deng M, Kwok TCY, Leung JCS, et al. All osteoporotically deformed vertebrae with >34% height loss have radiographically identifiable endplate/cortex fracture. *J Orthop Translat* 2018;14:63-6.
 21. Wáng YX, Che-Nordin N, Deng M, et al. Elderly males with or without existing osteoporotic vertebral fracture have much lower future vertebral fracture risk than elderly females: the MrOS (Hong Kong) year-4 follow-up spine radiograph study. *Osteoporos Int* 2019;30:2505-14.
 22. Wáng YX, Che-Nordin N. Informed communication with study subjects of radiographically detected osteoporotic vertebral deformity. *Quant Imaging Med Surg* 2018;8:876-80.
 23. Wáng YX, Santiago RF, Deng M, et al. Identifying osteoporotic vertebral endplate and cortex fractures. *Quant Imaging Med Surg* 2017;7:555-91.
 24. Wáng YX, Deng M, He LC, et al. Osteoporotic vertebral endplate and cortex fractures: a pictorial review. *J Orthop Translat* 2018;15:35-49.
 25. Lentle BC. Gender and the recognition of vertebral fractures. *Quant Imaging Med Surg* 2020;10:1401-7.
 26. Wáng YX, Wang XR, Che-Nordin N, et al. On the possibility of over-diagnosis of osteoporotic vertebral fracture at mid-thoracic level. *J Thorac Dis* 2019;11:5708-11.
 27. Leidig-Bruckner G, Limberg B, Felsenberg D, et al. Sex difference in the validity of vertebral deformities as an index of prevalent vertebral osteoporotic fractures: a population survey of older men and women. *Osteoporos Int* 2000;11:102-19.
 28. Kim YJ, Chae SU, Kim GD, et al. Radiographic Detection of Osteoporotic Vertebral Fracture without Collapse. *J Bone Metab* 2013;20:89-94.
 29. Eastell R, Cedel SL, Wahner HW, et al. Classification of vertebral fractures. *J Bone Miner Res* 1991;6:207-15.

30. McCloskey EV, Spector TD, Eyres KS, et al. The assessment of vertebral deformity: a method for use in population studies and clinical trials. *Osteoporos Int* 1993;3:138-47.
31. The International Osteoporosis Foundation (IOF): What We Do. Available online: <https://www.iofbonehealth.org/what-we-do/training-and-education/educational-slide-its/vertebral-fracture-teaching-program>, accessed on June 17, 2020.
32. Lentle B, Cheung AM, Hanley DA, et al. Scientific Advisory Council of Osteoporosis Canada. Osteoporosis Canada 2010 guidelines for the assessment of fracture risk. *Can Assoc Radiol J* 2011;62:243-50.
33. Johansson H, Oden A, McCloskey EV, et al. Mild morphometric vertebral fractures predict vertebral fractures but not non-vertebral fractures. *Osteoporos Int* 2014;25:235-41.
34. Wu CY, Li J, Jergas M, et al. Comparison of semiquantitative and quantitative techniques for the assessment of prevalent and incident vertebral fractures. *Osteoporos Int* 1995;5:354-70.
35. Buckens CE, de Jong PA, Mol C, et al. Intra and interobserver reliability and agreement of semiquantitative vertebral fracture assessment on chest computed tomography. *PLoS One* 2013;8:e71204.
36. Du MM, Che-Nordin N, Ye PP, et al. Underreporting characteristics of osteoporotic vertebral compressive fracture in back pain clinic patients of a tertiary hospital in China. *J Orthop Translat* 2020. doi: 10.1016/j.jot.2019.10.007.
37. Keen RW. Burden of osteoporosis and fractures. *Curr Osteoporos Rep* 2003;1:66-70.
38. Wáng YX. A modified semi-quantitative (mSQ) grading scheme for osteoporotic vertebral fracture in elderly women. *Quant Imaging Med Surg* 2019;9:146-50.
39. Wang XR, Xu FR, Huang QL, et al. Radiological features of traumatic vertebral endplate fracture: an analysis of 194 cases with 263 vertebral fractures. *Chin Med J (Engl)* 2020. [Epub ahead of print].

Cite this article as: Wáng YXJ, Liu WH, Diacinti D, Yang DW, Iannacone A, Wang XR, Kripa E, Che-Nordin N, Diacinti D. Diagnosis and grading of radiographic osteoporotic vertebral deformity by general radiologists after a brief self-learning period. *J Thorac Dis* 2020. doi: 10.21037/jtd-20-2379