Published in Journal of Clinical Densitometry 18, issue 2, 233-238, 2015
which should be used for any reference to this work

Reproducibility of Vertebral Fracture Assessment Readings From Dual-energy X-ray Absorptiometry in Both a Population-based and Clinical Cohort: Cohen's and Uniform Kappa

Bérengère Aubry-Rozier,[*,1] Roland Chapurlat,[2] François Duboeuf,[2] Katia Iglesias,[3] Marc-Antoine Krieg,[1] Olivier Lamy,[1] Bernard Burnand,[3] and Didier Hans[1]*

[1]*Centre for Bone Diseases, Lausanne University Hospital, Lausanne, Switzerland;*
[2]*Inserm UMR 1033, Hôpital Edouard Herriot, Lyon, France;*
[3]*Département de Médecine Sociale et Préventive, Lausanne University Hospital, Lausanne, Switzerland*

Abstract

Vertebral fracture assessments (VFAs) using dual-energy X-ray absorptiometry increase vertebral fracture detec-tion in clinical practice and are highly reproducible. Measures of reproducibility are dependent on the frequency and distribution of the event. The aim of this study was to compare 2 reproducibility measures, reliability and agreement, in VFA readings in both a population-based and a clinical cohort. We measured agreement and reliability by uniform kappa and Cohen's kappa for vertebral reading and fracture identification: 360 VFAs from a population-based cohort and 85 from a clinical cohort. In the population-based cohort, 12% of vertebrae were unreadable. Vertebral fracture prevalence ranged from 3% to 4%. Inter-reader and intrareader reliability with Cohen's kappa was fair to good (0.35−0.71 and 0.36−0.74, respectively), with good inter-reader and intrareader agreement by uniform kappa (0.74−0.98 and 0.76−0.99, respectively). In the clinical cohort, 15% of vertebrae were unreadable, and vertebral fracture prevalence ranged from 7.6% to 8.1%. Inter-reader reliability was moderate to good (0.43−0.71), and the agreement was good (0.68−0.91). In clinical situations, the levels of reproducibility measured by the 2 kappa statistics are concordant, so that either could be used to measure agreement and reliability. However, if events are rare, as in a population-based cohort, we recommend evaluating reproducibility using the uniform kappa, as Cohen's kappa may be less accurate.

Key Words: Cohen's kappa; population-based cohort; uniform kappa; vertebral fracture; VFA reproducibility.

Introduction

Osteoporotic fractures are a major cause of morbidity and mortality in industrialized countries *(1−3)*. Despite recent decreases in hip fracture incidence rates in industrialized countries, the incidence of other osteoporotic fractures is increasing, including symptomatic vertebral fractures (VFs) *(4)*. Nevertheless, VFs are underestimated and misdiagnosed in 50% of cases *(5)*. Moreover, knowledge of a VF can change the management of osteoporosis (in terms of treatment and follow-up).

Traditionally, the standard approach to identifying vertebral fractures has been through the acquisition of spinal radiographs. A specific semiquantitative analysis of VF, using Genant's grades *(6)*, is the classification system most often used in clinical practice and is highly reproduc-ible for the upper dorsal and lumbar spine *(7)*, with agree-ment ranging between 0.91 and 0.96 for prevalent VF. Over the past few years, however, a new imaging approach has been tested in clinical conditions, one that is associated with much less ionizing radiation (1% of the ionizing radi-ation dose associated with spinal radiographs) *(8)* and is coupled with bone mineral density (BMD) evaluation via the dual-energy

*Address correspondence to: Bérengère Aubry-Rozier, Center of Bone Diseases, Lausanne University Hospital, Avenue Pierre Decker 4, Lausanne 1011, Switzerland. E-mail: Berengere.aubry@chuv.ch

X-ray absorptiometry (DXA) examina-tion—this approach is called a vertebral fracture assessment (VFA). To validate this new examination, it was first neces-sary to establish its reproducibility relative to X-rays. Several teams have studied this reproducibility comparing spinal radiographs and VFA *(9–11)*. All authors have agreed that there is good reproducibility with ''highly read-able'' vertebrae, grades 2 and 3 by Genant's fracture defini-tion, and for the lumbar spine; but considerably worse results are obtained in other situations (grade 1 fractures and in the dorsal spine) *(10,12–16)*. Recently, the Interna-tional Osteoporosis Foundation and the International Soci-ety for Clinical Densitometry proposed performing VFA systematically coupled with DXA to detect VF and improve the utility of bone densitometry, thereby enhancing fracture risk assessment *(17,18)* (www.iofbonehealth.org). Follow-up plain-radiograph imaging is advisable if substantial numbers of vertebrae are not evaluable or if the presence of any deformity is uncertain (grade 1).

Most studies on VFA have been performed on osteoporosis patients in clinical settings, as opposed to population-based cohorts, to assess its use as a screening instrument. What is needed now is to test the reproducibility of the VFA in population-based cohort situations and perhaps adapt the guidelines. But can we use the same statistical test of repro-ducibility in a population-based cohort as in the clinic?

In all publications regarding the reproducibility of VFA, Cohen's kappa coefficient has been considered the most use-ful statistical test for qualitative data *(19)*. It is understood today that the question of reproducibility has 2 main compo-nents: reliability and agreement. But what do we measure with Cohen's kappa? Even in published studies, it is often not easy to know what we really want to measure and how. The kappa index is surrounded by controversy *(20)*. The main reason for this is a primary error in the design of any agreement study based on kappa when the sample is unrepre-sentative of the target population. A second problem occurs when the distribution of the event or condition of interest is nonuniform. Hence, interpretation of Cohen's kappa becomes inaccurate. This is why Cohen's kappa coefficient often is closer to a reliability than an agreement test *(21)*. To control for these frequent errors, a new test of agreement has recently been proposed: the uniform kappa *(20,22)*.

Compared with Cohen's kappa, the uniform kappa statistic captures agreement beyond hazards (a coin toss) and could be considered more of an agreement than reliability test.

Before conducting the present study, we hypothesized that in population-based cohorts, events are not uniformly distrib-uted and the reproducibility of VFA must be evaluated using the uniform kappa.

We aimed to compare the concordance between the 2 sta-tistical tests, uniform kappa for agreement and Cohen's kappa for reliability, in vertebral reading and vertebral fracture iden-tification by VFA in 2 populations with different prevalence rates of these events: the population-based cohort drawn from the OsteoLaus study and the clinical cohort selected from a clinic-based study conducted in Lyon *(23)*.

Materials and Methods

Study Population

The OsteoLaus cohort is a subpopulation of 1501 women, aged between 50 and 80 years, within the Lausanne cohort (CoLaus) *(24)*. The CoLaus study, which was a survey of roughly 6200 adults, was undertaken to gain a better under-standing of the association between cardiovascular disease and genetic factors in the general population of Lausanne, Switzerland. The principal aim of the OsteoLaus study has been to prospectively evaluate the bone status of a representa-tive sample of Lausanne's adult female population, by collect-ing data on clinical risk factors for osteoporotic fractures, as well as BMD evaluation by DXA, VFA, and trabecular bone score *(25)*. Rate of participation in the OsteoLaus study, started in July 2010, has been more than 85% than that of the CoLaus-selected patients. For the present study, we performed repro-ducibility analysis on 360 randomly selected OsteoLaus study patients to represent our population-based cohort.

The clinical study conducted in Lyon included 85 post-menopausal women undergoing BMD measurement either because of a painful vertebral fracture or because of some other reason for which BMD testing was indicated *(23)*.

VFA Within the Population-based Cohort

VFAs were performed for levels T4–L4 by means of lateral single-energy absorptiometry images of the thoracolumbar spine on Delphi A (Lyon) and Discovery A (Lausanne) densi-tometers (Hologic Inc., Bedford, MA). Images were acquired in the supine lateral position but with the patient in a poster-oanterior position. Only patients with a minimum of 5 analyz-able vertebrae were retained. If <5 vertebrae were analyzable, patients were classified as ''not interpretable'' and excluded from further analysis. Each reading, to determine if a VF was present or absent, was initially visual and qualitative. However, in cases of reader doubt, the semiquantitative method developed by Genant et al *(6)* was used directly on the computer screen, using a special software interface (Opta-sia Medical Viewer; SpineAnalyzer; Optasia Medical, Chea-dle, Cheshire, United Kingdom) that allowed for computerized measurement of specific vertebral heights. If the reader had no doubts about the presence or absence of VF, the semiquantitative method was not applied. We chose to largely classify our reading results as the following dichot-omous variables: (1) readable vertebrae, yes or no; and (2) on readable vertebrae only, vertebral fracture, yes or no. We also developed a 3-option ranking system that integrated nonread-able vertebrae; the 3 ranks were (1) not readable; (2) VF pre-sent; and (3) VF absent. All variables (related to reliability and agreement) were analyzed for total VFA and separately for VF in the dorsal spine and lumbar spine.

Expert Readings for the Population-based Cohort

Two independent readers (OL, BAR) read each of the 360 VFAs to assess inter-reading reproducibility. OL and BAR each have >10 years of experience in osteoporosis

management and regularly read VFA within their clinical practice. To assess intrareader reproducibility, BAR waited 5 weeks between each pair of readings so as to reduce the risk that she might remember her earlier interpretation. She also read all VFAs in a different order the second time through.

VFA Acquisition and Expert Readings for the Clinical Cohort

For the clinical cohort, 85 consecutive postmenopausal women undergoing BMD measurement as part of routine clinical practice were recruited. Single-energy 20-s morphometry scans (VFAs) were performed using both Hologic Delphi A and Discovery A densitometers. Images were acquired using decubitus lateral positioning (rotating C-arm) but with the patient again in a posteroanterior position. Two experts in VFA reading performed semiquantitative and visual qualitative assessments using the method developed by Genant et al, as previously published *(23)*, to allow for inter-reader analysis.

Statistical Analysis

To test inter-reader and intrareader reliability and agreement, we calculated Cohen's kappa and uniform kappa using the following formulas, respectively

$$\frac{Od/m - Ed/m}{1 - Ed/m}$$

where Od is the observed agreement, Ed is the expected agreement, and m represents total events

$$\frac{Od/m - 1/mod}{1 - 1/mod}$$

where m represents total events and mod represents number of modalities.

Depending on the formula, Cohen's kappa informs us about real variance proportional to global variance. Meanwhile, the uniform kappa coefficient assesses agreement between 2 objects, applying the same weight to every modality.

Kappa values, with 95% confidence intervals (CIs), were calculated for the dichotomous variables: readable vertebrae, yes or no; vertebral fracture, yes or no on readable vertebrae; no readable vertebrae or VF present or VF absent, for total VF, dorsal spine VF, and lumbar spine VF. All vertebrae were considered independently. We considered Landis and Koch's threshold values to interpret Cohen's kappa results. Traditionally, a Cohen's kappa value >0.81 is considered almost perfect; between 0.8 and 0.61 to be good; between 0.6 and 0.4 to be moderate; between 0.4 and 0.2 to be fair; between 0.2 and 0 to be slight; and <0 to be poor *(26)*. We considered any uniform kappa value >0.75 a good result *(22)*.

All statistical analyses were performed using the statistical software program Stata/IC12 (StataCorp LP, College Station, TX).

Results

For the overall population-based cohort (mean age 68 years; mean body mass index 25.7 kg/m^2), 12% of the vertebrae were unreadable, with this percentage as high as 48% for levels T4 and T5 (Fig. 1). Depending on the reader, the prevalence of VF varied from 3% to 4% (fracture or no fracture) for all vertebrae (4680) and from 17% to 24%among patients with at least 1 VF (Table 1). Inter-reader reli-ability with Cohen's kappa was fair to good (0.35−0.71; Table 2) for all variables. For the upper dorsal spine, the re-sults were worse (0.35−0.63). Inter-reader agreement beyond chance for the uniform kappa coefficient was good (0.74−0.98; Table 2) for all criteria. Intrareader reliability for Cohen's kappa also was fair to good (0.36−0.74) for all criteria, whereas uniform kappa's again were good, ranging from 0.76−0.99 (Table 3).

For the clinical cohort (mean age 71 years; mean body mass index 24 kg/m^2), 15% of the vertebrae were unreadable, with this percentage as high as 77% and 47% for levels T4 and T5, respectively (Fig. 2). Depending on the reader, the prevalence of VF varied from 7.6% to 8.1% (fracture or no fracture) across all vertebrae (1105; Table 1). Inter-reader reliability with Cohen's kappa was moderate to good (0.43−0.71), whereas inter-reader agreement beyond chance for the uniform kappa coefficient was good (0.68−0.91) for all variables (Table 4). The levels of reproducibility measured in the clinical cohort were more comparable than in the population-based cohort, in which they were quite discordant (Table 4).

Discussion

In the population-based cohort, we found that results for Cohen's kappa were fair to good, as per Landis and Koch's rankings *(26)*, indicating a moderate level of reproducibility, but the uniform kappa was consistently high, indicating a high level of reproducibility. Reproducibility results using the 2 tests were discordant. In the clinical study cohort *(23)*, we found moderate-to-good reliability for Cohen's

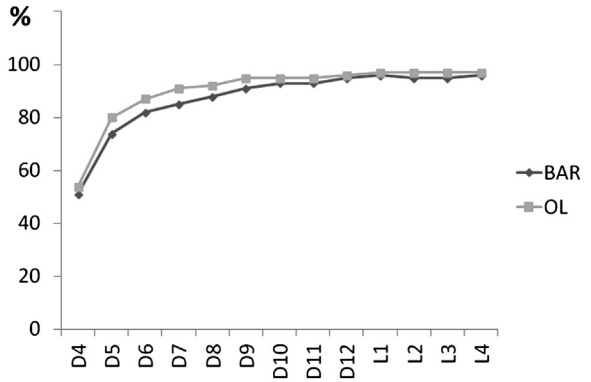

Fig. 1. Readability. Comparison between the 2 readers (BAR and OL).

Table 1
Patient Data

OsteoLaus	Reader 1 number of patients	Reader 2 number of patients	Lyon	Reader 1 number of patients	Reader 2 number of patients
Without fracture	286	255	Without fracture	44	41
With at least 1 VF	61	88	With at least 1 VF	41	44
With 2 VF	13	13	With 2 VF	10	12
With 3 VF	2	3	With 3 VF	6	4
With 4 VF	0	1	With 4 VF	2	3
With >4 VF	0	0	With >4 VF	6	6

Note: Total number of patients assessed: 360 from OsteoLaus and 85 from Lyon. Total number of vertebrae assessed: 4680 (3240 dorsal and 1440 lumbar).

Abbr: VF, vertebral fracture.

kappa and good agreement with uniform kappa. The reproducibility results were concordant with the 2 tests.

For both clinical and economic considerations, early and accurate recognition of vertebral fractures is very important. Clinically, this is true because the diagnosis of a prevalent or an incident VF is both important and a challenge. This diagnosis may cause changes both in the monitoring and treatment of the disease. Economically, this is true because of how reimbursements are made for different treatments. The gold standard to diagnose a VF is spinal X-rays; however, the International Society for Clinical Densitometry and International Osteoporosis Foundation have recently proposed that VFA be performed, in addition to a DXA examination, so as to enhance fracture risk assessment. VFA images are sufficiently reproducible in well-controlled clinical practice for patients suffering from osteoporosis with either prevalent or incident VF *(7,9−16)*. Only a small number of studies have been published analytically comparing 2 VFA and/or repeated VFA. Buehring et al *(16)* identified better VFA readings using Lunar iDXA than Prodigy (General Electric; Milwaukee, WI), and better reproducibility when VFAs were read by an experienced reader. In addition to this, the patient position used during the acquisition of lateral VFA images may influence reproducibility *(27,28)*. Unfortunately, the use of VFA as a screening instrument is less well documented. To justify using the VFA method in clinical and population-based populations, one must have acceptable familiarity with the tool, as well as some appreciation regarding its advantages and limits. Reproducibility of readings is a crucial component of this familiarity.

To date, Cohen's kappa has been the favored test in scientific research when estimating reproducibility; however, it tends to be more of a reliability than agreement test *(21)*.

Table 2
VFA in the OsteoLaus Cohort: Inter-reader Results

Criteria	Cohen's kappa (95% CI)	Kappa uniform (95% CI)
Readable, all spine	0.61 (0.58−0.65)	0.84 (0.82−0.85)
Readable, dorsal spine	0.63 (0.59−0.67)	0.81 (0.79−0.83)
Readable, lumbar spine	0.40 (0.29−0.51)	0.90 (0.88−0.92)
VF if readable, all spine	0.47 (0.38−0.56)	0.95 (0.94−0.96)
VF if readable, dorsal spine	0.41 (0.30−0.51)	0.93 (0.92−0.95)
VF if readable, lumbar spine	0.71 (0.55−0.88)	0.98 (0.97−0.99)
Ranking, all spine	0.58 (0.58−0.59)	0.85 (0.84−0.86)
Ranking, dorsal spine	0.59 (0.56−0.61)	0.82 (0.80−0.83)
Ranking, lumbar spine	0.47 (0.45−0.51)	0.91 (0.89−0.93)

Note: Total number of vertebrae assessed: 4680 (3240 dorsal and 1440 lumbar).

Abbr: CI, confidence interval; VF, vertebral fracture; VFA, vertebral fracture assessment.

Table 3
VFA in OsteoLaus Cohort: Intrareader Results

Criteria	Cohen's kappa (95% CI)	Uniform kappa (95% CI)
Readable, all spine	0.74 (0.71−0.77)	0.89 (0.88−0.90)
Readable, dorsal spine	0.73 (0.70−0.77)	0.86 (0.85−0.88)
Readable, lumbar spine	0.69 (0.59−0.79)	0.95 (0.94−0.97)
VF if readable, all spine	0.41 (0.31−0.51)	0.95 (0.94−0.96)
VF if readable, dorsal spine	0.36 (0.25−0.47)	0.94 (0.93−0.95)
VF if readable, lumbar spine	0.60 (0.40−0.80)	0.98 (0.96−0.99)
Ranking, all spine	0.69 (0.65−0.70)	0.89 (0.88−0.90)
Ranking, dorsal spine	0.68 (0.66−0.69)	0.86 (0.85−0.87)
Ranking, lumbar spine	0.66 (0.61−0.68)	0.95 (0.94−0.96)

Note: Total number of vertebrae assessed: 4680 (3240 dorsal, 1440 lumbar).

Abbr: CI, confidence interval; VF, vertebral fracture.

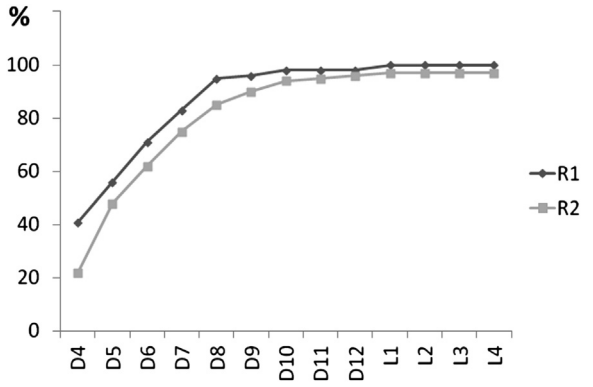

Fig. 2. Lyon cohort readability. Comparison between the 2 readers (R1 and R2).

To better evaluate agreement, the uniform kappa is proposed by most statisticians *(20,21)*.

A small number of clinical studies have already demonstrated how agreement can be calculated using the uniform kappa rather than Cohen's kappa, especially if the event of interest is rare. In a recently published study *(29)* evaluating self-reports of diabetes care by patients and physicians, authors highlighted how important it is to clearly distinguish between reliability and agreement in medical studies. They identified discordances between agreement (calculated using the uniform kappa), which tended to be high, and reliability (calculated using Cohen's kappa), which tended to be low, as tests of reproducibility for the dichotomous indicators in their study. They explained this apparent discordance by noting asymmetry in the marginal distributions of the obser-vations. Finally, they concluded that, in terms of reproduc-ibility, high reliability may be preferable to discriminate between patients, but high agreement may be better to estab-lish a diagnosis. Until our present study, when VFA readings were evaluated in a clinical cohort, the level of reproducibility calculated using Cohen's kappa was precise enough to yield reliable results in terms of diagnosis. Nevertheless, when we wanted to assess the reproducibility of VFA readings in a population-based cohort, Cohen's kappa seemed inadequate due to the asymmetry in marginal distributions that exists because of the small number of events. To maintain credibility of the VFA examination, and to obtain or maintain reimbursement for its use, researchers, clinicians, and publishers need to use a more accurate statistical test to analyze their findings. With increased familiarity regarding how each statistical test is defined and best interpreted, readers, colleagues and public health care services may be better able to understand the additional value of the VFA examination. In this article, we wanted to highlight this point using a concrete example of 2 distinct patient populations, one a population-based cohort and the other a clinical cohort in which the prevalence of VF was virtually doubled. We found concordant results when reproducibility was evaluated with either kappa in the clinical cohort but discordant results between the 2 tests in the population-based cohort. This, we surmise, will be especially true when the event of interest is rare and nonasymmetrically distributed. For example, in the population-based cohort, there was uniform distribution in terms of all vertebrae being readable; correspondingly, the difference between Cohen's (0.61) and uniform Kappa (0.84) was relatively small. On the other hand, when the event was highly asymmetrical—for example, vertebral fractures—the 2 kappa values were very different (0.47 vs 0.95).

One main limitation of our study is the absence of conventional radiography of the spine and any comparison against VFA. To minimize radiation exposure and costs, X-ray evaluations have not been part of the OsteoLaus study. A second limitation is that our results are applicable only in situations in which reliability and agreement are tested among expert readers of VFA. To address this issue, another study is currently in progress to test (1) if reproducibility is affected in a population-based cohort by a reader's level of expertise; and (2) if, in this situation, the degree of discordance between the 2 statistical tests is less relative to nonexpert readers.

We conclude that, in clinical situations in which an event is frequent, reproducibility estimates for VFA will typically be accurate and concordant using either kappa value. On the other hand, if events are rare or nonsymmetrically distributed—for example, within population-based cohorts—the 2 kappa results will be discordant because Cohen's kappa tends to underestimate the true level of reproducibility. In this latter case, we recommend using the uniform kappa.

Table 4
Cohen and Uniform Kappa Estimates in OsteoLaus and Lyon Cohorts

Criteria	OsteoLaus Cohen (95% CI)	OsteoLaus uniform (95% CI)	Lyon Cohen (95% CI)	Lyon uniform (95% CI)
Readable, all spine	0.74 (0.71−0.77)	0.89 (0.88−0.90)	0.62 (0.56−0.67)	0.76 (0.72−0.79)
Readable, dorsal spine	0.73 (0.70−0.77)	0.86 (0.85−0.88)	0.61 (0.55−0.66)	0.68 (0.63−0.73)
Readable, lumbar spine	0.69 (0.59−0.79)	0.95 (0.94−0.97)	0.43 (0.14−0.72)	0.76 (0.62−0.88)
VF if readable, all spine	0.41 (0.31−0.51)	0.95 (0.94−0.96)	0.69 (0.62−0.77)	0.89 (0.86−0.92)
VF if readable, dorsal spine	0.36 (0.25−0.47)	0.94 (0.93−0.95)	0.68 (0.58−0.79)	0.89 (0.84−0.92)
VF if readable, lumbar spine	0.60 (0.40−0.80)	0.98 (0.96−0.99)	0.71 (0.60−0.83)	0.90 (0.85−0.94)

Abbr: CI, confidence interval; VF, vertebral fracture.

References

1. Lippuner K, Golder M, Greiner R. 2005 Epidemiology and direct medical costs of osteoporotic fractures in men and women in Switzerland. Osteoporos Int 16(2 Suppl):S8−S17.
2. Wustrack R, Seeman E, Bucci-Rechtweg C, et al. 2012 Predic-tors of new and severe vertebral fractures: results from the HO-RIZON Pivotal Fracture Trial. Osteoporos Int 23(1):53−58.
3. Cauley JA. 2013 Public health impact of osteoporosis. J Geron-tol A Biol Sci Med Sci 68(10):1243−1251.
4. Lippuner K, Popp AW, Schwab P, et al. 2011 Fracture hospital-izations between years 2000 and 2007 in Switzerland: a trend analysis. Osteoporos Int 22(9):2487−2497.
5. Casez P, Uebelhart B, Gaspoz JM, et al. 2006 Targeted education improves the very low recognition of vertebral fractures and osteoporosis management by general internists. Osteoporos Int 17(7): 965−970.
6. Genant HK, Wu CY, van Kuijk C, Nevitt MC. 1993 Vertebral fracture assessment using a semi-quantitative technique. J Bone Miner Res 8(9):1137−1148.
7. Wu C, van Kuijk C, Li J, et al. 2000 Comparison of digitized images with original radiography for semiquantitative assessment of osteoporotic fractures. Osteoporos Int 11(1):25−30.
8. Damilakis J, Adams JE, Guglielmi G, Link TM. 2010 Radiation exposure in X-ray-based imaging techniques used in osteoporosis. Eur Radiol 20(11):2707−2714.
9. 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 Miner Res 15(3):575−585.
10. Schousboe JT, Debold CR. 2006 Reliability and accuracy of vertebral fracture assessment with densitometry compared to radiography in clinical practice. Osteoporos Int 17(2):281−289.
11. Ferrar L, Jiang G, Schousboe JT, et al. 2008 Algorithm-based qual-itative and semiquantitative identification of prevalent vertebral fracture: agreement between different readers, imaging modal-ities, and diagnostic approaches. J Bone Miner Res 23(3): 417−424.
12. Binkley N, Krueger D, Gangnon R, et al. 2005 Lateral vertebral assessment: a valuable technique to detect clinically significant vertebral fractures. Osteoporos Int 16(12):1513−1518.
13. Rea JA, Li J, Blake GM, et al. 2000 Visual assessment of vertebral deformity by X-ray absorptiometry: a highly predictive method to exclude vertebral deformity. Osteoporos Int 11(8): 660−668.
14. Ferrar L, Jiang G, Clowes JA, et al. 2008 Comparison of densito-metric and radiographic vertebral fracture assessment using the algorithm-based qualitative (ABQ) method in postmenopausal women at low and high risk of fracture. J Bone Miner Res 23(1): 103−111.
15. Grados F, Fechtenbaum J, Flipon E, et al. 2009 Radiographic methods for evaluating osteoporotic vertebral fractures. Joint Bone Spine 76(3):241−247.
16. Buehring B, Krueger D, Checovich M, et al. 2010 Vertebral fracture assessment: impact of instrument and reader. Osteoporos Int 21(3):487−494.
17. Rosen HN, Vokes TJ, Malabanan AO, et al. 2013 The Official Positions of the International Society for Clinical Densitom-etry: vertebral fracture assessment. J Clin Densitom 16(4): 482−488.
18. Schousboe JT, Vokes T, Broy SB, et al. 2008 Vertebral Fracture Assessment: the 2007 ISCD Official Positions. J Clin Densitom 11(1):92−108.
19. Cohen J. 1960 A coefficient of agreement for nominal scales. Educ Psychol Measure 20:37−46.
20. Erdmann T, De Mast J, Warrens M. 2012 Some common errors of experimental design, interpretation and inference in agreement studies. Stat Methods Med Res. http://dx.doi.org/10.1177/0962280211433597. [Epub ahead of print].
21. de Vet HC, Terwee CB, Knol DL, Bouter LM. 2006When to use agreement versus reliability measures. J Clin Epidemiol 59(10):1033−1039.
22. de Mast J. 2007 Agreement and kappa-type indices. Am Stat 61(2):148−153.
23. Chapurlat RD, Duboeuf F, Marion-Audibert HO, et al. 2006 Effectiveness of instant vertebral assessment to detect prevalent vertebral fracture. Osteoporos Int 17(8):1189−1195.
24. Firmann M, Mayor V, Vidal PM, et al. 2008 The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. BMC Cardiovasc Disord 8:6.
25. Silva BC, Leslie WD, Resch H, et al. 2014 Trabecular bone score: a noninvasive analytical method based upon the DXA image. J Bone Miner Res 29(3):518−530.
26. Landis JR, Koch GG. 1977 An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 33(2):363−374.
27. Pearson D, Horton B, Green DJ, et al. 2006 Vertebral morphometry by DXA: a comparison of supine lateral and decubitus lateral densitometers. J Clin Densitom 9(3):295−301.
28. Paggiosi MA, Finigan J, Peel N, et al. 2012 Supine vs. decubitus lateral patient positioning in vertebral fracture assessment. J Clin Densitom 15(4):454−460.
29. Collet TH, Taffe P, Bordet J, et al. 2014 Reproducibility of dia-betes quality of care indicators as reported by patients and phy-sicians. Eur J Public Health 24(6):1004−1009.