

Comparison of expert and nonexpert physicians in the assessment of vertebral fractures using the semiquantitative method in Japan

Yukari Uemura · Nobuaki Miyakawa · Hajime Orimo ·
Masataka Shiraki · Toshitaka Nakamura · Satoshi Mori

Received: 28 January 2014 / Accepted: 5 August 2014 / Published online: 10 October 2014
© The Japanese Society for Bone and Mineral Research and Springer Japan 2014

Abstract Assessment of vertebral fracture is critically important for the diagnosis and treatment of osteoporosis. This study aimed to clarify the effectiveness of the semi-quantitative (SQ) method in the assessment of vertebral fractures in Japanese clinical practice. Forty-four physicians (seven experts and 37 nonexperts) assessed the spinal radiographs of 40 patients participating in the Adequate Treatment of Osteoporosis (A-TOP) Japanese Osteoporosis Intervention Trial (JOINT)-02 at the baseline, 12 months, and 24 months using the SQ method. The proportion of diagnosed fracture cases per spine was higher in the non-expert group than in the expert group at each time point, and was especially high in the upper thoracic spine (T4–T6). The least mean squares spinal fracture index was significantly higher in the nonexpert group than in the expert

group for all time points. The kappa statistics were also higher in the expert group than in the nonexpert group for all vertebral levels at all time points. Assessment of vertebral fractures using the SQ method tended to be overestimated by nonexpert physicians compared with the experts, with poor nonexpert interobserver reliability and well-matched expert interobserver reliability. Conscious efforts to avoid overestimation and to obtain higher reliability with the SQ method should be made to achieve more precise diagnoses and treatment of osteoporosis in Japanese clinical practice.

Keywords Osteoporosis · Semiquantitative method · Spinal fracture index · Vertebral fracture

Y. Uemura
Department of Biostatistics, School of Public Health,
The University of Tokyo, Tokyo, Japan

N. Miyakawa
Public Health Research Foundation, Tokyo, Japan

H. Orimo
Japan Osteoporosis Foundation, Tokyo, Japan

M. Shiraki
Department of Internal Medicine, Research Institute and Practice
for Involutional Diseases, Nagano, Japan

T. Nakamura
National Center for Global Health and Medicine, Tokyo, Japan

S. Mori (✉)
Bone and Joint Surgery, Seirei Hamamatsu General Hospital,
2-12-12 Sumiyoshi, Naka-ku, Hamamatsu, Shizuoka 430-8558,
Japan
e-mail: stmori@sis.seirei.or.jp

Introduction

Vertebral fractures are the commonest osteoporotic fractures, and the assessment of vertebral fractures is widely used to diagnose osteoporosis or monitor disease progression. This assessment has also been used in many clinical trials as an end point to evaluate the efficacy of drugs for the treatment of osteoporosis [1–4]. Several methods of assessing vertebral fractures have been developed, and most are categorized as quantitative morphometry (QM) [5–9]. The morphometric approach is based on a comparison between the vertebral heights of osteoporotic patients and the vertebral height of normal women, including the anterior–posterior ratio, middle–posterior ratio, and posterior–posterior adjacent ratio. The cutoff thresholds differ, and no single measurement is considered the gold standard for vertebral fracture assessment. Recently, the semiquantitative (SQ) method has been used to assess vertebral fractures in clinical practice and clinical trials instead of QM.

Genant et al. [10] devised the SQ method as a new way to assess vertebral fractures without measuring vertebral heights. In the SQ method, each spine was graded into four categories (normal, mild, moderate, and severe) on visual inspection. Excellent reproducibility of interobserver and intraobserver reliability (between experienced and inexperienced but trained observers) was found, with good agreement between QM and SQ methods [10]. Wu et al. [11] found excellent interobserver agreement using the SQ method. Grados et al. [12] compared the SQ method with four morphometric methods for assessing prevalent vertebral fractures, and found good agreement as well. Crans et al. [13] revealed that a spinal deformity index derived from the SQ assessment of vertebral fractures predicted future vertebral fracture risk and that the spinal deformity index or SQ method was clinically useful for the treatment of osteoporosis.

Despite these studies, little is known of exactly how the SQ method is used in clinical practice in Japan. Therefore, our aim was to clarify how the SQ method is used to assess vertebral fractures in clinical practice in Japan, by comparing expert physicians with nonexpert physicians.

Materials and methods

Materials

Lateral (thoracic and lumbar) spine radiographs of 40 osteoporotic patients were included in the present study. All of the radiographs originated from the Japanese Osteoporosis Intervention Trial (JOINT)-02 conducted by the Adequate Treatment of Osteoporosis (A-TOP) research group to evaluate combination therapy (alendronate and alfacalcidol) compared with monotherapy (alendronate alone) nationwide in Japan. The details of JOINT-02, including the study design, patient characteristics, inclusion and exclusion criteria, and end points, were previously reported [14, 15]. Baseline and follow-up (12 and 24 months) radiographs were converted into electronic data files (Digital Imaging and Communication in Medicine files). The use of the radiographs in this study was approved by the A-TOP executive and the ethical committees.

SQ method and spinal fracture index

The SQ approach was developed by Genant et al. [10] in the 1990s as a new method to assess vertebral fractures on visual inspection without measuring vertebral heights. The grading of each spine is classified into four categories as follows: normal (grade 0); mild deformity (grade 1, 20–25 % reduction in anterior, middle, and/or posterior height and 10–20 % reduction in area); moderate deformity

(grade 2, 25–40 % reduction in any height and 20–40 % reduction in area); and severe deformity (grade 3, 40 % reduction in any height and area). In each spine, grade 1 or higher was considered “fractured” and grade 0 was considered “not fractured.” The spinal fracture index (SFI) was calculated for each patient by dividing the sum of individual vertebral grade scores by the number of spines evaluated, which provided general information on the osteoporosis severity in an individual patient [10].

Assessment of vertebral fractures using the SQ method

Seven expert physicians (expert group) and 37 nonexpert physicians (nonexpert group) independently assessed the vertebral deformity grade (T4–L4) of each patient on a personal computer using the SQ method. The expert group consisted of three orthopedists, three spinal surgeons, and one radiologist (average medical career, 29 years), all with experience assessing vertebral fractures in several drug trials or highly specialized experience assessing vertebral fractures in clinical practice. The nonexpert group consisted of 18 orthopedists, 14 internal medicine physicians, and 5 radiologists (average medical career, 16 years), not using the SQ method for assessing vertebral fractures in daily practice. Baseline and follow-up radiographs were assessed in chronological order per patient. The physician assessment data were gathered and statistically analyzed as a data set.

Statistical analysis

The frequency and proportion of SQ grade per spine and per visit were assessed between the experts and nonexperts, and the proportions were compared using the Pearson chi squared test. Also, the proportion of radiographs that the experts and nonexperts assessed as fractured (grade 1 or higher) was examined per spine at the baseline, 12 months, and 24 months. Physicians assessed the SFI for each patient, and the mean value was calculated. Using the SFI as the dependent variable, we used a mixed effects model, accounting for the correlation between the SFI of each patient assessed by the same physician. The SFI least mean squares was estimated to evaluate the difference between the experts and the nonexperts using the model, adjusted and not adjusted for their years of medical experience. To analyze interobserver reproducibility within each group, we calculated the kappa statistics per group per spine (T4–L4). Because we were interested in the degree of agreement between more than two physicians, we used the extended kappa statistical method proposed by Fleiss [16]. The kappa groups were as follows: 0–0.2, poor agreement; 0.2–0.4, fair agreement; 0.4–0.6, moderate agreement; 0.6–0.8, good agreement; 0.8–1.0, very good agreement [16].

Table 1 Proportion of semiquantitative grade per spine level assessed by the expert and nonexpert groups

	Grade	Baseline			12 months			24 months		
		Expert group (%)	Nonexpert group (%)	<i>P</i>	Expert group (%)	Nonexpert group (%)	<i>P</i>	Expert group (%)	Nonexpert group (%)	<i>P</i>
T4	0	95.4	71.3	<0.0001	93.6	67.4	<0.0001	91.8	64.5	<0.0001
	1	3.9	22.2		3.6	22.7		5.4	24.2	
	2	0.0	3.4		1.1	4.4		1.1	6.0	
	3	0.0	0.1		1.4	2.3		1.8	2.5	
	ND	0.7	3.0		0.4	3.2		0.0	2.8	
T5	0	91.8	73.1	<0.0001	91.8	71.2	<0.0001	91.1	67.4	<0.0001
	1	6.1	21.0		6.4	23.4		6.8	25.3	
	2	1.8	4.0		1.8	4.6		2.1	6.1	
	3	0.0	0.0		0.0	0.1		0.0	0.1	
	ND	0.4	1.9		0.0	0.7		0.0	1.1	
T6	0	92.5	76.0	<0.0001	91.4	72.2	<0.0001	90.4	68.9	<0.0001
	1	5.7	18.7		6.4	22.8		6.8	25.1	
	2	1.8	3.8		2.1	4.6		2.9	5.3	
	3	0.0	0.0		0.0	0.1		0.0	0.3	
	ND	0.0	1.5		0.0	0.3		0.0	0.5	
T7	0	82.9	68.2	<0.0001	80.0	63.2	<0.0001	78.6	59.2	<0.0001
	1	8.9	20.7		8.9	23.8		10.0	25.9	
	2	4.3	6.7		4.6	7.4		3.6	7.0	
	3	3.9	3.0		6.4	5.3		7.9	7.4	
	ND	0.0	1.4		0.0	0.2		0.0	0.5	
T8	0	88.2	71.0	<0.0001	83.6	69.2	<0.0001	81.1	64.6	<0.0001
	1	5.0	18.7		7.5	20.6		7.1	22.0	
	2	3.2	5.8		4.6	5.6		5.0	6.3	
	3	3.6	3.5		4.3	4.3		6.8	6.6	
	ND	0.0	1.0		0.0	0.3		0.0	0.5	
T9	0	90.0	70.9	<0.0001	85.4	66.8	<0.0001	80.4	60.9	<0.0001
	1	3.2	18.8		4.3	19.9		4.3	21.2	
	2	4.3	6.9		4.3	7.2		6.8	9.3	
	3	2.5	2.5		6.1	5.4		8.6	8.2	
	ND	0.0	1.0		0.0	0.7		0.0	0.3	
T10	0	81.1	66.4	<0.0001	80.0	63.4	<0.0001	77.1	59.9	<0.0001
	1	11.1	22.4		11.8	24.0		12.1	26.0	
	2	3.2	5.3		3.6	6.9		6.1	8.3	
	3	4.6	5.1		4.6	4.9		4.6	5.4	
	ND	0.0	0.9		0.0	0.9		0.0	0.4	
T11	0	74.6	57.4	<0.0001	67.5	50.1	<0.0001	62.9	46.9	<0.0001
	1	13.9	28.6		14.6	30.9		14.6	30.1	
	2	8.2	9.0		13.9	13.2		15.0	13.3	
	3	3.2	3.9		3.6	4.4		6.4	8.3	
	ND	0.0	1.2		0.4	1.5		1.1	1.5	
T12	0	51.1	37.2	<0.0001	46.1	32.0	<0.0001	43.9	28.9	<0.0001
	1	12.1	23.4		12.1	23.2		10.7	23.4	
	2	18.6	19.1		18.9	19.7		17.9	18.2	
	3	18.2	20.1		22.9	24.5		27.1	28.9	
	ND	0.0	0.3		0.0	0.5		0.4	0.7	

Table 1 continued

	Grade	Baseline			12 months			24 months		
		Expert group (%)	Nonexpert group (%)	<i>P</i>	Expert group (%)	Nonexpert group (%)	<i>P</i>	Expert group (%)	Nonexpert group (%)	<i>P</i>
L1	0	53.9	41.6	<0.0001	46.4	34.2	<0.0001	41.8	29.9	<0.0001
	1	13.9	24.4		15.0	24.7		12.1	21.4	
	2	19.3	17.2		21.1	18.7		23.9	22.4	
	3	12.9	16.3		17.5	21.8		22.1	26.1	
	ND	0.0	0.5		0.0	0.7		0.0	0.3	
L2	0	78.9	60.2	<0.0001	71.4	50.7	<0.0001	66.1	45.3	<0.0001
	1	5.7	23.6		6.8	25.2		5.7	23.5	
	2	9.3	8.8		11.8	12.6		16.8	16.0	
	3	6.1	7.3		10.0	11.5		11.4	15.1	
	ND	0.0	0.1		0.0	0.0		0.0	0.1	
L3	0	90.4	67.8	<0.0001	75.0	56.0	<0.0001	65.4	46.1	<0.0001
	1	7.5	25.6		7.5	26.1		9.3	26.2	
	2	2.1	6.1		17.1	14.2		21.4	18.4	
	3	0.0	0.5		0.4	3.5		3.9	9.0	
	ND	0.0	0.1		0.0	0.3		0.0	0.3	
L4	0	85.0	65.8	<0.0001	72.5	56.0	<0.0001	69.6	50.5	<0.0001
	1	5.7	21.0		7.5	21.4		7.9	22.8	
	2	6.8	7.2		15.4	14.1		14.3	15.1	
	3	1.8	4.3		3.6	6.9		5.7	8.6	
	ND	0.7	1.6		1.1	1.7		2.5	3.1	

Null hypothesis; equal to the proportion of semiquantitative grade evaluated between the expert group and nonexpert group in each spine level
ND not determined

Results

Forty-four physicians (seven experts and 37 nonexperts) assessed 40 sets of spine radiographs at the baseline and during the follow-up period using the SQ method. Table 1 shows the proportions of SQ grade per spine level assessed by the expert and nonexpert groups. There was a significant difference in all spine levels and at all time points between the two groups. The proportion of grade 0 was lower for every spine level in the nonexpert group than in the expert group. Figure 1 shows the proportion of fractured cases per spine level that the expert and nonexpert physicians assessed at the baseline, 12 months, and 24 months. The proportion of fractured cases was high in the thoracolumbar spine (T11–L2) compared with other spine levels in both groups. In addition, the proportion per spine was higher in the nonexpert group than in the expert group at each time point, and was especially high in the upper thoracic spine (T4–T6).

The mean values of the SFI assessed per case by the experts and nonexperts were plotted at the baseline, 12 months, and 24 months (Fig. 2). The mean values were consistently higher in the nonexpert group than in the expert group for every time point.

We compared the SFI of the expert group with that of the nonexpert group at the baseline, 12 months, and 24 months using a mixed effects model adjusted or not adjusted for years of experience as a physician. The least mean squares SFI was significantly higher in the nonexpert group than in the expert group for all time points ($P < 0.0001$) (Fig. 3). The margins of the least mean squares SFI between the nonexpert group and the expert group remained almost constant regardless of adjustment, at 0.21 (not adjusted) and 0.19 (adjusted) at the baseline, 0.21 and 0.19 at 12 months, and 0.23 and 0.19 at 24 months, respectively.

Table 2 shows the interobserver kappa statistics in the expert and nonexpert groups for SQ grade of vertebral deformity per spine. The kappa statistics were higher in the expert group than in the nonexpert group for all vertebral levels at the baseline, 12 months, and 24 months. The expert group scores were considered to have moderate or good agreement, except for the T4 and L3 levels at the baseline, and the T5, and T6 levels at 24 months, and were particularly high between T12 and L4. The kappa statistics for the nonexpert group were considered to have poor or fair agreement at the baseline, 12 months, and 24 months, and were particularly low between T4 and T6.

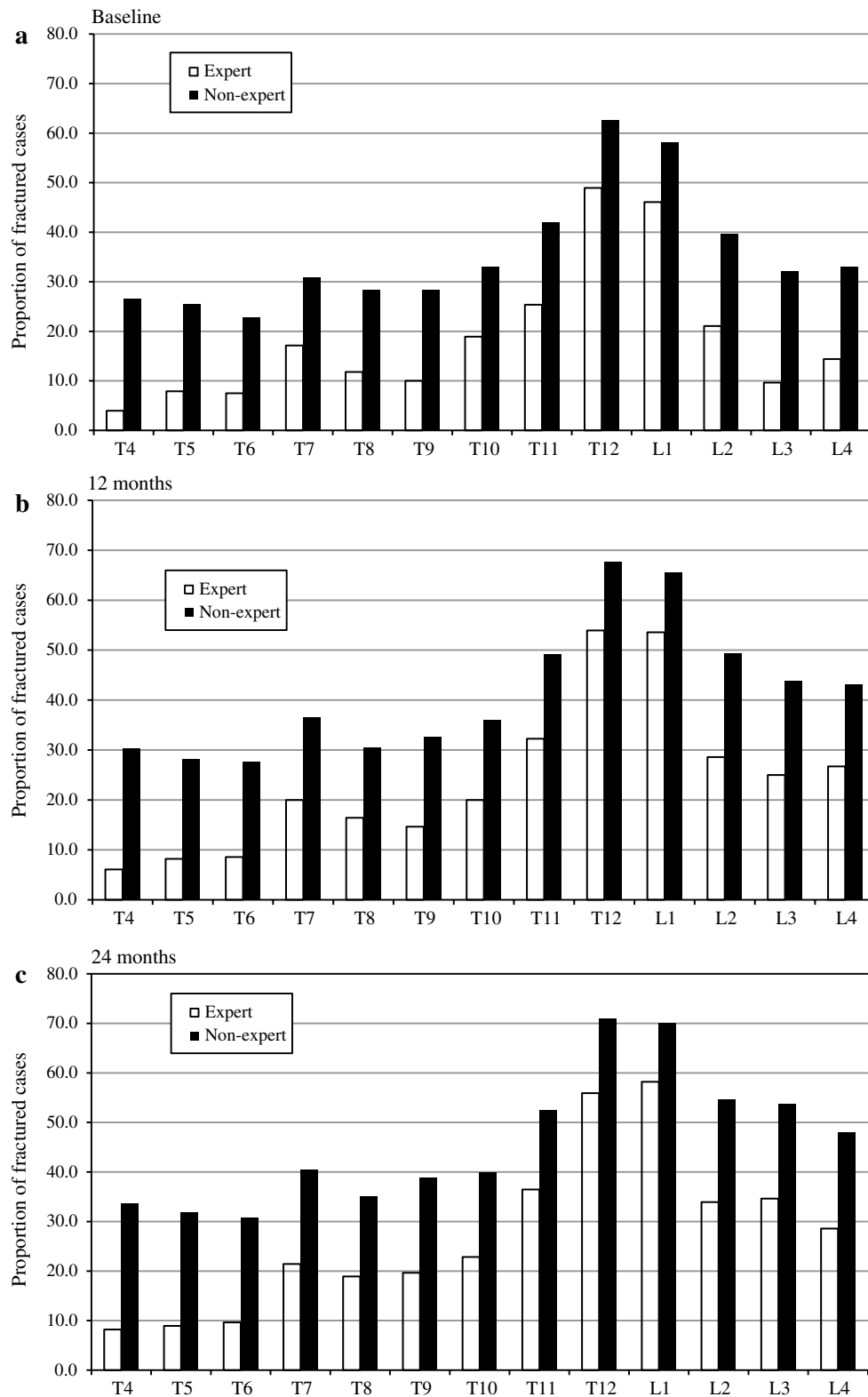


Fig. 1 Proportion of fractured cases assessed by the expert and nonexpert groups: **a** baseline, **b** 12 months, **c** 24 months

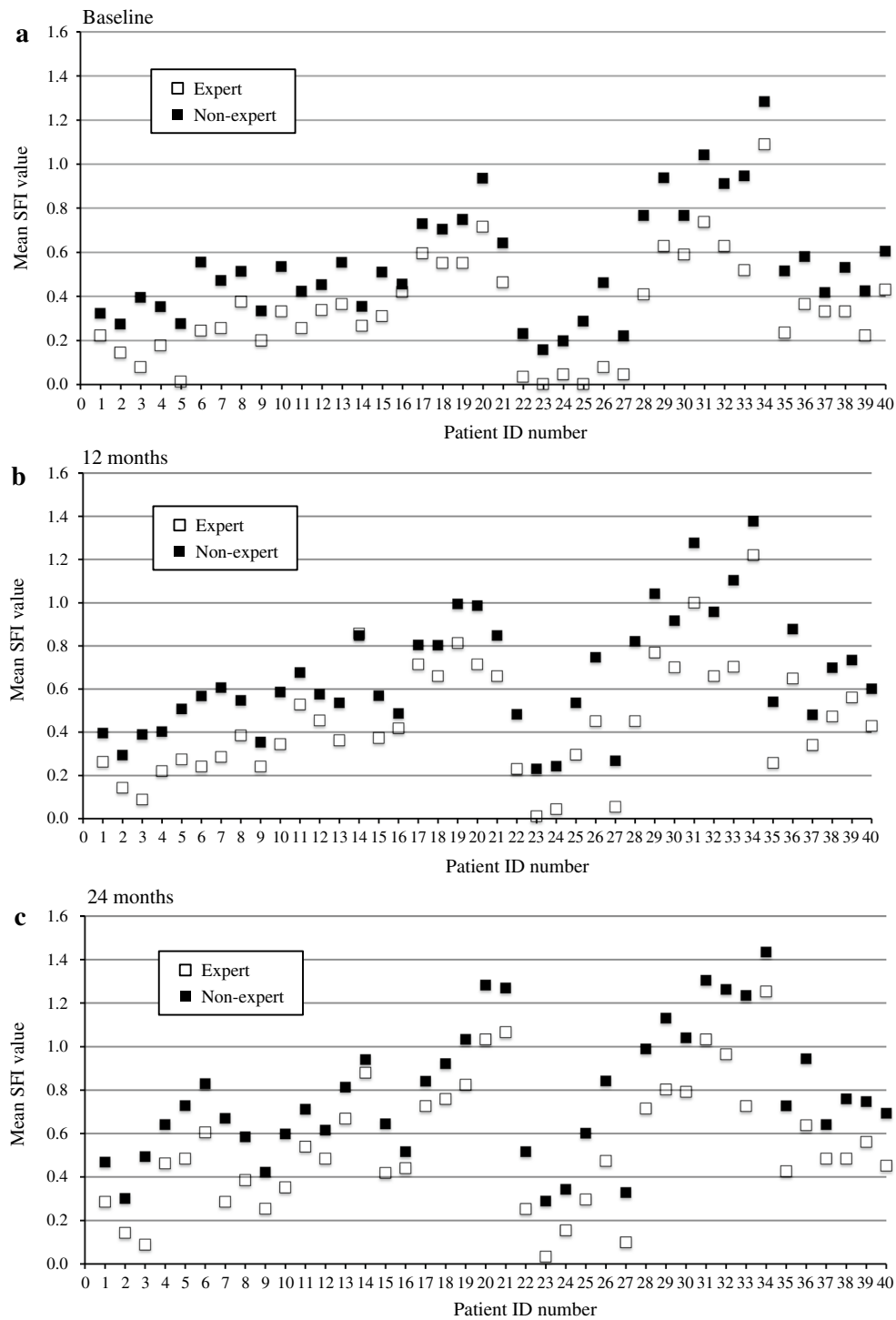


Fig. 2 Mean spinal fracture index (SFI) per patient assessed by the expert and nonexpert groups: **a** baseline, **b** 12 months, **c** 24 months

Discussion

We assessed the vertebral fractures of 40 patients using the SQ method at the baseline, 12 months, and 24 months, and evaluated the interobserver reproducibility and discrepancies between expert and nonexpert physicians. In all spines

from T4 to L4, the proportion of fractured cases (grade 1 or higher) was higher in the nonexpert group than in the expert group at the baseline and in the follow-up period (Fig. 1), and the proportion of SQ grades evaluated was significantly different per spine between the two groups (Table 1). In addition, in all cases, the mean value of the SFI was

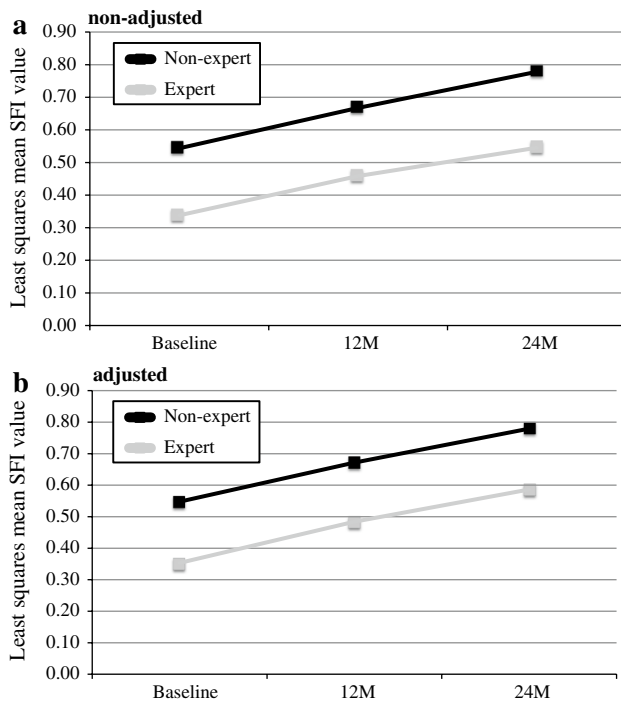


Fig. 3 Expert versus nonexpert comparison of SFI from the baseline to 24 months: **a** not adjusted, **b** adjusted

higher in the nonexpert group than in the expert group at all visits (Fig. 2). The nonexpert group had a tendency to overestimate the SQ grade of vertebral fractures, particularly in the thoracic spine, compared with the experts. Genant et al. [10, 11] reported that the SQ assessment of vertebral fractures showed an excellent intraobserver and interobserver agreement between experienced and inexperienced but trained physicians, and that the SQ method was a good reproducible method to assess osteoporotic vertebral

fractures. However, another report indicated that it was difficult to identify subtle differences between SQ grade 1 as mild fracture and borderline deformity (grade 0.5), and those assessments were sometimes arbitrary [17]. In these reports, the inexperienced physicians were well trained [10, 17], and it would appear that they understood how to make use of all information regarding vertebral body size, shape, and projection to assess vertebral fractures using the SQ method. The discrepancies between the two groups in our study may have resulted from a lack of previous training in SQ assessment for the nonexperts. In addition, our study was conducted under several preexisting biases because the radiographs we used originated from JOINT-02, and it was previously reported that participants in that study had high fracture risks, with the number of prevalent vertebral fractures of one or more as an inclusion criterion [13, 14], which may have had some effect on the assessment of vertebral fractures by the nonexpert group.

The least mean squares SFI was significantly higher in the nonexpert group than in the expert group at the baseline and in the follow-up period, whether adjusted or not adjusted (Fig. 3). The estimated margin between the two groups was fairly constant at 0.19–0.23. These results also indicate an overestimation by the nonexpert group compared with the expert group for the SQ assessment of vertebral fractures. Conversely, there was a major difference between the expert group and the nonexpert group in the kappa statistics at the baseline and in the follow-up period. The kappa statistics for the nonexpert group were notably low at 0–0.2 (poor agreement) from T4 to T6 and 0.2–0.4 (fair agreement) from T7 to L4, whereas those of the expert group were high at 0.4–0.6 (moderate agreement) in most spine levels and 0.6–0.8 (good agreement) from T12 to L4. The interobserver reproducibility in the expert group for SQ assessment was excellent compared with that of the

Table 2 Kappa statistics per spine level in both groups at the baseline and in the follow-up period

	Baseline		12 months		24 months	
	Expert group	Nonexpert group	Expert group	Nonexpert group	Expert group	Nonexpert group
T4	0.243	0.120	<i>0.447</i>	0.146	<i>0.498</i>	0.152
T5	<i>0.498</i>	0.136	<i>0.471</i>	0.115	0.396	0.103
T6	<i>0.450</i>	0.120	<i>0.433</i>	0.141	0.358	0.127
T7	<i>0.570</i>	0.261	<i>0.604</i>	0.293	<i>0.606</i>	0.291
T8	<i>0.484</i>	0.222	<i>0.600</i>	0.254	<i>0.654</i>	0.287
T9	<i>0.515</i>	0.219	<i>0.527</i>	0.267	<i>0.633</i>	0.302
T10	<i>0.582</i>	0.276	<i>0.545</i>	0.289	<i>0.554</i>	0.285
T11	<i>0.479</i>	0.210	<i>0.538</i>	0.238	<i>0.555</i>	0.256
T12	<i>0.648</i>	0.363	<i>0.647</i>	0.358	<i>0.664</i>	0.365
L1	<i>0.654</i>	0.397	<i>0.640</i>	0.385	<i>0.620</i>	0.392
L2	<i>0.713</i>	0.313	<i>0.757</i>	0.363	<i>0.709</i>	0.367
L3	0.383	0.163	<i>0.704</i>	0.255	<i>0.629</i>	0.310
L4	<i>0.606</i>	0.248	<i>0.619</i>	0.295	<i>0.627</i>	0.304

Italic font indicates kappa scores of moderate to good agreement; *normal font* indicates poor to fair agreement

nonexpert group, similar to the findings of Genant et al. [10, 11].

Delmas et al. [18] reported that underdiagnosis of vertebral fractures was observed in the IMPACT trial (a multicenter multinational prospective study) in several geographic regions, including North America, Latin America, Europe, South Africa, and Australia. All radiologists were given a radiographic procedure manual, which was the principal tool for standardization of the SQ assessment, and this was a major difference compared with our study. The results of Delmas et al.'s study were as follows: there were 789 patients with vertebral fractures (grade 1 or higher) and 1,662 patients without vertebral fractures (grade 0) in the central readings, and 607 with vertebral fractures and 1,844 without vertebral fractures in the local readings. Further, among 789 patients with vertebral fractures in the central readings, 266 patients had no vertebral fractures (false-negative rate, 34 %) and 523 patients had vertebral fractures (true-positive rate, 66 %) in the local readings. Among 1,662 patients without vertebral fractures in the central readings, 1,578 patients had no vertebral fractures (true-negative rate, 95 %) and 84 patients had vertebral fractures (false-positive rate, 5 %) in the local readings. Conversely, our results indicated that the proportion of fractured cases was lower in the expert group than in the nonexpert group, revealing a discrepancy in the results between the two studies. It appears that a bias toward aggressive identification of vertebral fractures occurs in a clinical trial because of the strict protocol and use of the radiographic procedure manual, which differed from our study design, and our results may be reasonable despite no use of a manual.

Our study has several limitations. First, the SQ assessment was performed using images on a personal computer rather than on the actual X-ray films, which may have reduced the image resolution and minimized the shape of the spine. Second, all physicians independently assessed the vertebral fractures from the baseline to 24 months without instructions for the standardizing of the SQ assessment method, such as would be obtained from a special manual. Finally, the results of our expert group are not a gold standard of assessment but a reference, and our results are a relative comparison, only, of the two groups because there was no assessment adjudication in the expert group.

Vertebral fracture assessment is important not only in diagnosis and evaluation of the treatment effects of osteoporosis but also in epidemiologic studies of osteoporosis or the treatment of the clinical vertebral fracture. The SQ method may be not well known in daily clinical practice, but it has been widely used in assessment of vertebral fracture in many clinical trials of osteoporotic drugs. Precise assessment of vertebral fracture using the SQ method in daily practice is necessary to realize proper diagnosis and treatment of osteoporosis. Our results suggests that (1)

conscious effort should be made to promote the SQ method in daily practice, and (2) training programs for the SQ method may be helpful to avoid overestimation of vertebral fractures by nonexpert physicians.

In conclusion, we clarified that the SQ assessment of vertebral fractures tended to be overestimated by nonexpert physicians, with poor nonexpert interobserver reliability and well-matched expert physician interobserver reliability in Japan. The SQ method is generally understood to include the entire spectrum of features of spinal deformity and to have a high reproducibility. Conscious efforts should be made to promote the SQ method to contribute to the treatment of osteoporosis.

Acknowledgments The authors thank the expert and nonexpert physicians for the assessment of the vertebral fractures using the SQ method. We also thank the A-TOP research group for providing the spinal radiographs of the patients participating in JOINT-02.

Conflict of interest All authors have no conflicts of interest.

References

1. Chesnut CH III, Skag A, Christiansen C et al (2004) Effects of oral ibandronate administered daily or intermittently on fracture risk in postmenopausal osteoporosis. *J Bone Miner Res* 19:1241–1249
2. Harris ST, Watts NB, Genant HK et al (1999) Effects of risedronate treatment on vertebral and nonvertebral fractures in women with postmenopausal osteoporosis: a randomized controlled trial. *JAMA* 282:1344–1352
3. Ettinger B, Black DM, Mitlak BH et al (1999) Reduction of vertebral fracture risk in postmenopausal women with osteoporosis treated with raloxifene: results from a 3-year randomized clinical trial. *JAMA* 282:637–645
4. Neer RM, Arnaud CD, Zanchetta JR et al (2001) Effect of parathyroid hormone (1–34) on fractures and bone mineral density in postmenopausal women with osteoporosis. *N Engl J Med* 344:1434–1441
5. Melton LJ 3rd, Lane AW, Cooper C et al (1993) Prevalence and incidence of vertebral deformities. *Osteoporos Int* 3:113–119
6. Eastell R, Cedel SL, Wahner HW et al (1991) Classification of vertebral fractures. *J Bone Miner Res* 6:207–215
7. Minne HW, Leidig G, Wüster C et al (1988) A newly developed spine deformity index (SDI) to quantitate vertebral crush fractures in patients with osteoporosis. *Bone Miner* 3: 335–349
8. Sauer P, Leidig G, Minne HW et al (1991) Spine deformity index (SDI) versus other objective procedures of vertebral fracture identification in patients with osteoporosis: a comparative study. *J Bone Miner Res* 6:227–238
9. McCloskey EV, Spector TD, Eyres KS et al (1993) The assessment of vertebral deformity: a method for use in population studies and clinical trials. *Osteoporos Int* 3:138–147
10. Genant HK, Wu CY, van Kuijk C et al (1993) Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res* 8:1137–1148
11. Wu CY, Li J, Jergas M et al (1995) Comparison of semiquantitative and quantitative techniques for the assessment of prevalent and incident vertebral fractures. *Osteoporos Int* 5:354–370

12. Grados F, Roux C, de Vernejoul MC et al (2001) Comparison of four morphometric definitions and a semiquantitative consensus reading for assessing prevalent vertebral fractures. *Osteoporos Int* 12:716–722
13. Crans GG, Genant HK, Krege JH et al (2005) Prognostic utility of a semiquantitative spinal deformity index. *Bone* 37:175–179
14. Shiraki M, Kuroda T, Miyakawa N et al (2011) Design of a pragmatic approach to evaluate the effectiveness of concurrent treatment for the prevention of osteoporotic fractures: rationale, aims and organization of a Japanese Osteoporosis Intervention Trial (JOINT) initiated by the Research Group of Adequate Treatment of Osteoporosis (A-TOP). *J Bone Miner Metab* 29:37–43
15. Orimo H, Nakamura T, Fukunaga M et al (2011) Effects of alendronate plus alfacalcidol in osteoporosis patients with a high risk of fracture: the Japanese Osteoporosis Intervention Trial (JOINT)-02. *Curr Med Res Opin* 27:1273–1284
16. Fleiss JL (1971) Measuring nominal scale agreement among many raters. *Psychol Bull* 76:378–382
17. Genant HK, Jergas M, Palermo L et al (1996) Comparison of semiquantitative visual and quantitative morphometric assessment of prevalent and incident vertebral fractures in osteoporosis. *J Bone Miner Res* 11:984–996
18. Delmas PD, van de Langerijt L, Watts NB et al (2005) Underdiagnosis of vertebral fractures is a worldwide problem: the IMPACT study. *J Bone Miner Res* 20:557–563