

A Call for Consensus Regarding Focal Structural Defects

KE Hansen, N Binkley, MK Drezner, D Malone, R Christian, D Krueger, N Vallarta-Ast, RD Blank

University of Wisconsin Osteoporosis Research Program and Clinical Center, Madison, Wisconsin

BACKGROUND

The International Society for Clinical Densitometry (ISCD) has published criteria by which to exclude vertebrae from lumbar spine densitometric analysis [1]. When applying these exclusion criteria there is only moderate agreement between interpreters, with poorest agreement regarding the identification of focal structural defects (FSD) [2]. To further investigate the utility of the ISCD vertebral body exclusion criteria, we assessed *intra*-observer reproducibility and specifically whether interpreters identify FSD consistently on two separate occasions (Part 1). We found that both intra- and inter-observer agreement regarding FSD was poor. We subsequently initiated group meetings to achieve consensus regarding FSD of sufficient severity to merit vertebral body exclusion. Once we achieved group consensus, inter-observer agreement was tested again (Part 2) and was substantially improved.

METHODS, PART I

Three ISCD-certified technologists obtained the scans in a precision study; subjects included 64 women and 26 men with a mean age, weight and L1-L4 T-score of 59.1 ± 9.3 years, 71.7 ± 13.8 kg and 0.1 ± 1.2, respectively. Five ISCD-certified physicians applied the ISCD vertebral exclusion criteria to ninety de-identified lumbar spine DXA printouts on two occasions at least four weeks apart. A standardized worksheet was used to record the vertebrae excluded, indications for exclusion and final T-scores. We had no access to our initial notes when reviewing bone density studies on the second occasion. We analyzed resulting data where appropriate with descriptive, Pearson's correlation coefficient (R), and kappa (κ) test statistics.

κ is the ability to identically apply the ISCD vertebral body exclusion criteria by one interpreter on two occasions, or by two interpreters on one occasion. Cut points for κ are:

- κ ≤ 0, poor agreement
- 0 < κ ≤ 0.2, slight agreement
- 0.2 < κ ≤ 0.4, fair agreement
- 0.4 < κ ≤ 0.6, moderate agreement
- 0.6 < κ ≤ 0.8, substantial agreement
- κ > 0.8, excellent agreement

RESULTS, PART I

When applying the ISCD exclusion criteria other than FSD, intra-observer κ were generally good or excellent, with over 60% of values falling between 0.6 and 1.0 (Table 1). Additionally, intra-observer T-score reproducibility was outstanding, with very high correlation coefficients (R = 0.97-0.99) and small absolute differences in final lumbar spine T-score between paired analyses (Table 2).

In contrast, focal structural defects were a source of both intra- and inter-observer disagreement. Specifically, half of intra-observer κ values were <0.60, indicating only moderate or fair agreement. Moreover, when determining the presence of FSD, enormous inter-observer variability was also observed. For example, the five physicians excluded L4 in 23, 28, 41, 51 and 89% of 90 scans (Figure 1). There was no relationship between κ and percent of

vertebrae excluded by each interpreter (R= -0.13, p=0.59), nor could we demonstrate a relationship between each interpreter's absolute change in T-score between paired reports and his or her rate of vertebral body exclusion (R=0.22, p=0.72).

Table 1 Intra-Observer Kappa Test Statistics with Application of ISCD Exclusion Criteria

Vertebra	Interpreter	Exclude	FSD	UND	BA	BMC
1	1	0.76 (0.09)	0.71 (0.12)	1.0	0.78 (0.12)	0.90 (0.10)
	2	0.82 (0.08)	0.82 (0.08)	0.66 (0.32)	1.0	0.50 (0.31)
	3	0.71 (0.10)	0.89 (0.08)	0.38 (0.28)	0.31 (0.25)	0.26 (0.23)
	4	0.51 (0.15)	0.54 (0.16)	0.64 (0.16)	-0.22 (0.01)	-0.02 (0.02)
	5	1	1	0.66 (0.32)	0.66 (0.32)	0.38 (0.28)
2	1	0.63 (0.10)	0.58 (0.11)	0.64 (0.17)	0.68 (0.14)	0.31 (0.25)
	2	0.57 (0.09)	0.47 (0.09)	0.64 (0.17)	0.96 (0.04)	0.71 (0.16)
	3	0.74 (0.08)	0.71 (0.10)	0.59 (0.12)	0.77 (0.10)	0.29 (0.18)
	4	0.77 (0.10)	0.54 (0.16)	0.78 (0.11)	0.48 (0.22)	0.22 (0.21)
	5	1	1	0.82 (0.12)	0.39 (0.27)	1
3	1	0.71 (0.08)	0.53 (0.10)	0.85 (0.06)	0.88 (0.12)	0.65 (0.23)
	2	0.36 (0.11)	0.32 (0.11)	0.75 (0.08)	0.93 (0.07)	0.71 (0.16)
	3	0.75 (0.07)	0.62 (0.10)	0.76 (0.08)	0.59 (0.15)	0.59 (0.15)
	4	0.73 (0.08)	0.47 (0.15)	0.78 (0.07)	0.55 (0.23)	-0.03 (0.02)
	5	0.95 (0.03)	0.95 (0.03)	0.94 (0.04)	0.38 (0.28)	0.75 (0.14)
4	1	0.73 (0.07)	0.63 (0.08)	0.82 (0.09)	0.90 (0.07)	0.96 (0.04)
	2	0.55 (0.12)	0.5 (0.12)	0.61 (0.12)	1.0	0.87 (0.07)
	3	0.67 (0.08)	0.52 (0.10)	0.72 (0.09)	0.73 (0.11)	0.62 (0.12)
	4	0.55 (0.09)	0.40 (0.12)	0.78 (0.08)	0.45 (0.16)	0.47 (0.15)
	5	0.98 (0.02)	0.98 (0.02)	0.82 (0.12)	1	1

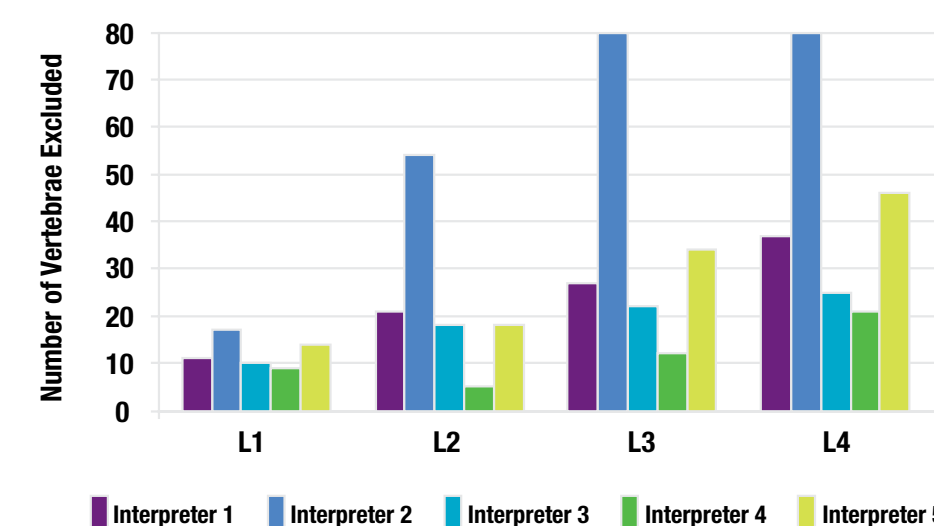
The standard deviation of each kappa test is shown in parentheses. "FSD" denotes focal structural defects, "UND" is an unusual discrepancy in T-score between adjacent vertebrae, and "BMC" and "BA" indicate a lack of increase in bone mineral content and area when progressing caudally from L1 to L4.

Table 2 Test-Retest Reproducibility of Final T-score with Application of Exclusion Criteria

Interpreter	Absolute Δ T-score (± Standard deviation)	Pearson's correlation coefficient (95% confidence interval)
1	0.072 ± 0.192	0.99 (0.98-0.99)
2	0.135 ± 0.266	0.97 (0.95-0.98)
3	0.067 ± 0.148	0.99 (0.99-1.0)
4	0.140 ± 0.364	0.97 (0.95-0.98)
5	0.017 ± 0.126	0.99 (0.99-1.00)

Figure 1 Inter-Observer Disagreement Regarding Vertebral Body Exclusion for Focal Structural Defects

Figure 1 shows, among 90 scans, the absolute number of vertebrae excluded by each interpreter.



METHODS, PART II

Marked high inter-observer disagreement regarding the presence of focal structural defects demanded subsequent group consensus regarding those lumbar spine artifacts of sufficient severity to warrant vertebral body exclusion. To accomplish this, we identified patients in our own clinics with paired bone density tests and lumbar spine radiographs performed within the same six months (n=20). Individually we reviewed the de-identified bone density tests to determine vertebrae affected by FSD. Then as a group, we met to identify sources of disagreements regarding whether to exclude a vertebra. To resolve these disagreements, the bone density image and lumbar spine radiograph were reviewed for each patient, and a group consensus reached regarding characteristics that merited vertebral body exclusion or inclusion.

Based on group discussions, we developed an annotated atlas of vertebral focal structural defects. To test the ability of the atlas to improve inter-observer agreement regarding FSD, two interpreters reviewed the 90 original bone density tests again, applying the exclusion criteria as described above. Subsequently, inter-observer κ between two interpreters was substantially improved (Figure 2).

Figure 2 A Comparison of Inter-Observer Agreement Before and After Consensus Regarding Focal Structural Defects

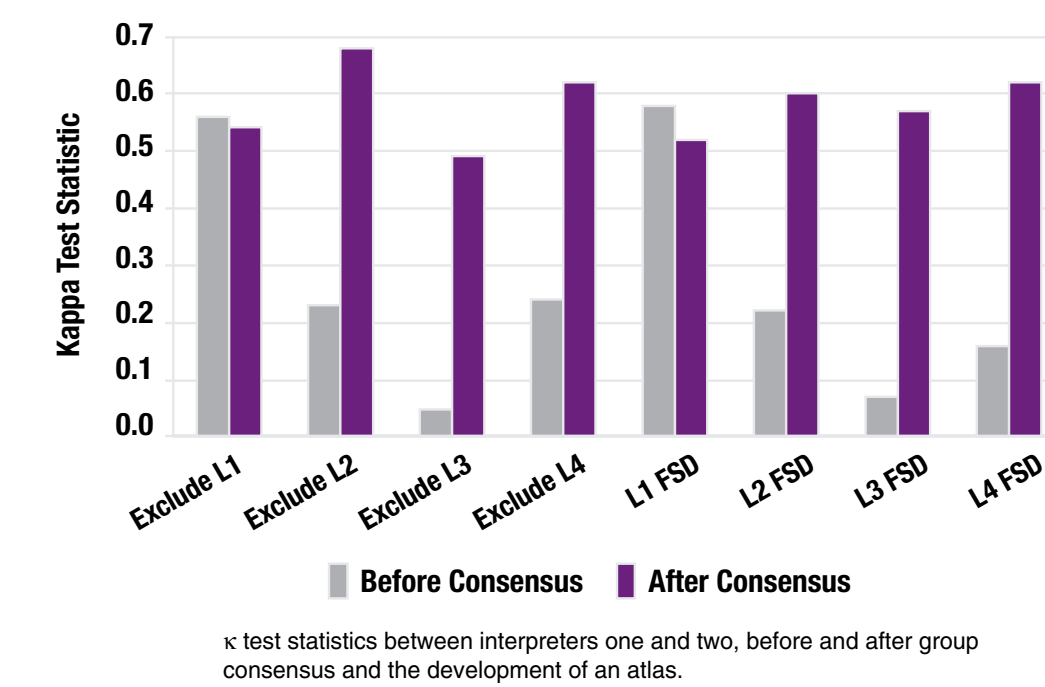
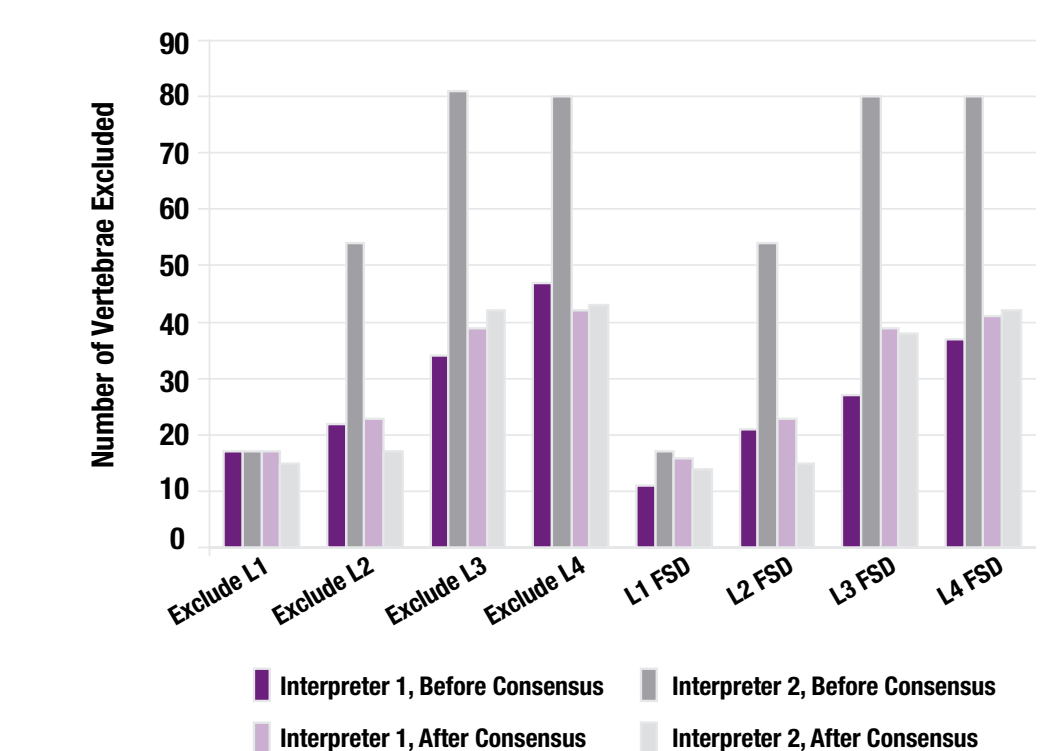
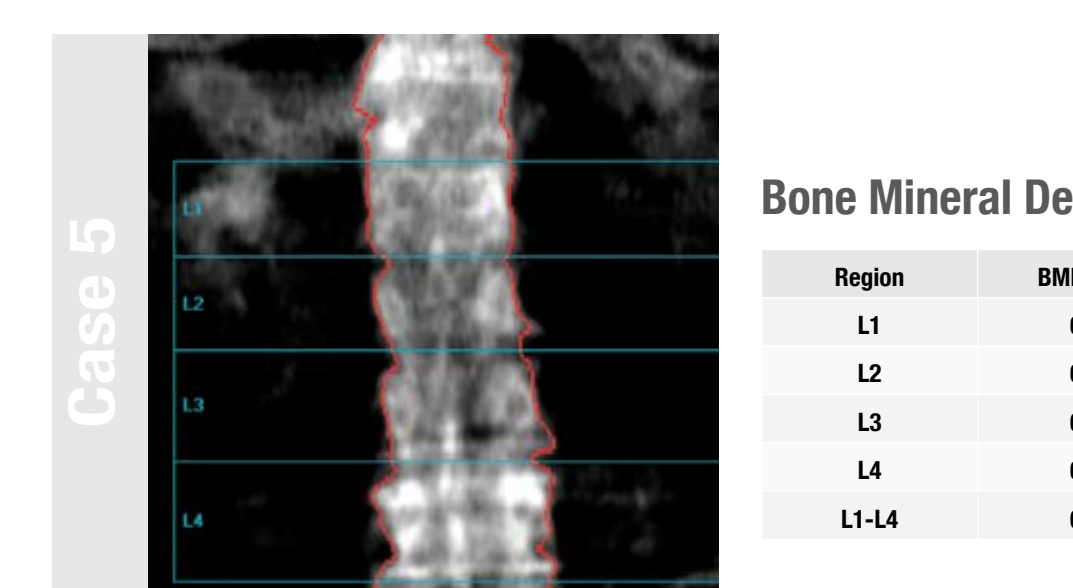
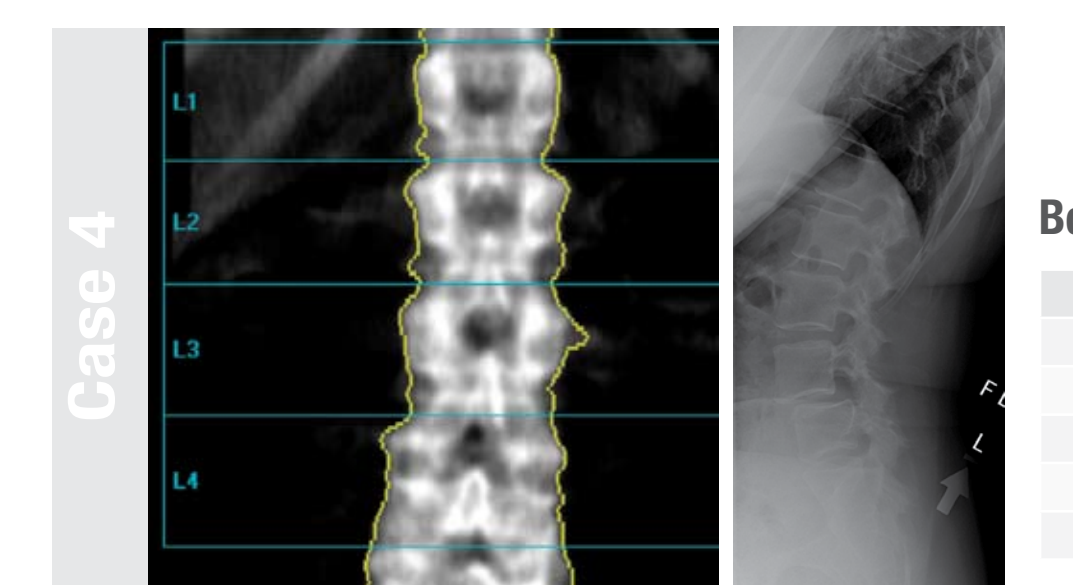
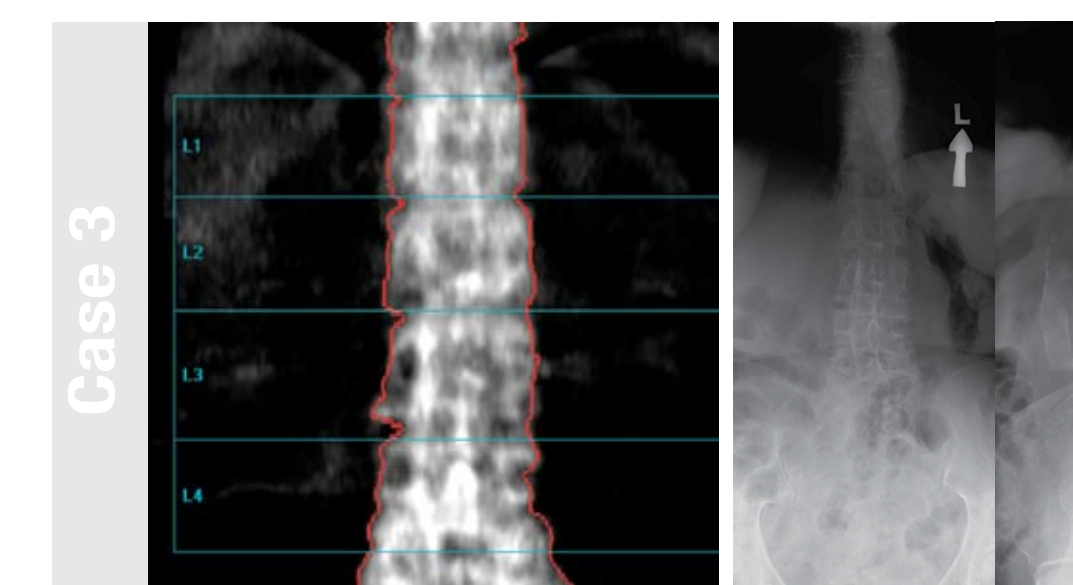
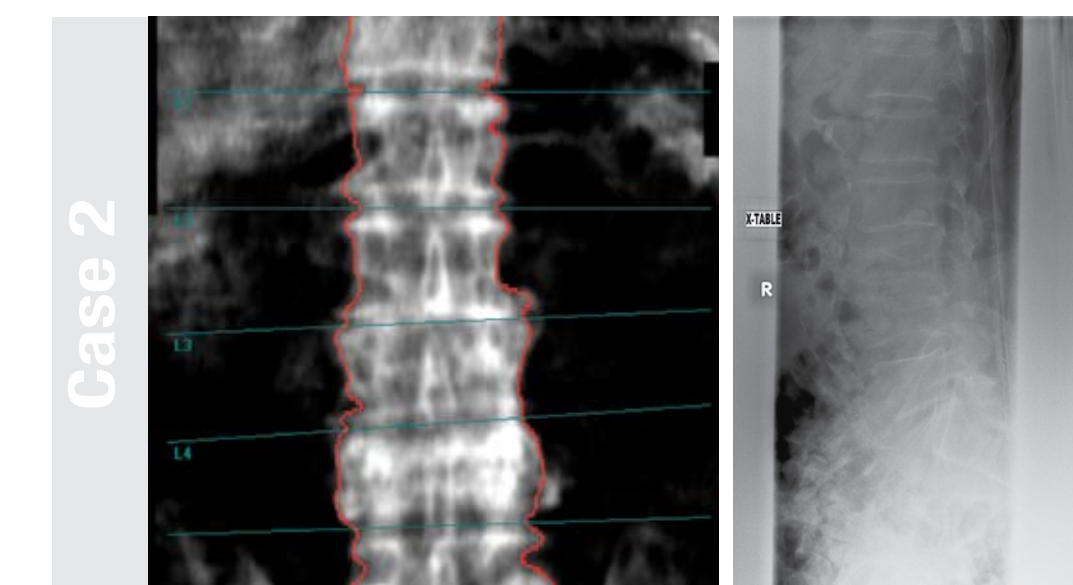
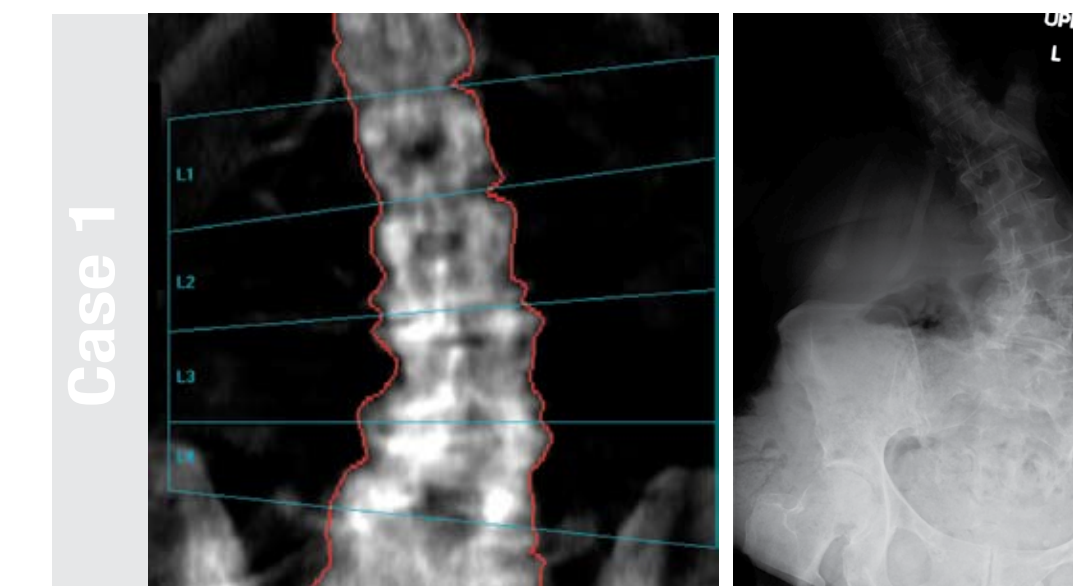


Figure 3: Differences in Rate of Vertebral Body Exclusion before and after Consensus, for Two Interpreters



Examples from Focal Structural Defect Atlas and Legend

- A vertebral focal structural defect accompanied by an unusual discrepancy in T-score in comparison to adjacent vertebrae should prompt exclusion of that vertebral body from the final lumbar spine T-score.
- Vertebrae should be suspected of having FSD if any of the following characteristics are present:
 - Asymmetry in image density between left and right sides of the vertebra (Case 1, Exclude L2, L3 and L4)
 - Osteophytes that substantially alter image density (Case 2, Exclude L2, L3 and L4)
 - A narrow horizontal strip of high density at the top of a vertebral body may indicate a compression fracture, as confirmed by x-ray (Case 2, L2 and L3)
 - Calcification that expands beyond one vertebra, suggesting aortic calcification, as confirmed by x-ray (Case 3)
- An interpreter may retain vertebra affected by focal structural defects, if inclusion of the vertebra provides valuable clinical information regarding lumbar spine bone mass. However, in this situation the interpreter should comment that a defect is present. Such vertebrae may be included within the final T-score when:
 - There is no T-score discrepancy between adjacent vertebrae (Case 4)
 - Only minor asymmetry in image density is noted (Case 4)
 - Very low T-scores are present despite focal abnormalities, providing useful clinical information regarding the presence of low bone mass (Case 5)



CONCLUSIONS

In conclusion, interpreters demonstrate high intra-observer T-score reproducibility when applying the ISCD exclusion criteria on two occasions at least four weeks apart. However, both intra- and inter-observer disagreement exist regarding focal structural defects; such disagreement potentially impacts on patient diagnosis and treatment. Since FSD are a major source of lower inter-observer and intra-observer agreement, standardized definitions exist regarding what constitutes an FSD should in principle improve the quality and consistency of lumbar spine DXA interpretation.

We conclude that group consensus and the subsequent development of an atlas improves agreement between interpreters regarding vertebral body exclusion and focal structural defects. An annotated atlas may prove to be a useful reference for clinicians who regularly interpret bone density studies.

REFERENCE

- Lenchik, L., et al., Executive summary International Society for Clinical Densitometry position development conference Denver, Colorado July 20-22, 2001. J Clin Densitom, 2002. 5 Suppl. p. S1-3.
- Hansen, K.E., et al., Interobserver Reproducibility of Criteria for Vertebral Body Exclusion. J Bone Miner Res, 2005. 20(3): p. 501-508.

Bone Mineral Density Data, Case 4

Region	BMD, g/cm ²	T-score
L1	1.087	-0.1
L2	1.119	-0.7
L3	1.186	-0.1
L4	1.188	-0.1
L1-L4	1.145	-0.3

Bone Mineral Density Data, Case 5

Region	BMD, g/cm ²	T-score
L1	0.635	-4.1
L2	0.539	-5.5
L3	0.571	-5.2
L4	0.793	-3.4
L1-L4	0.653	-4.4