Interobserver Agreement Using Computed Tomography to Assess Radiographic Fusion Criteria With a Unique Titanium Interbody Device

Paul J. Slosar, MD, Jay Kaiser, MD, Luis Marrero, MD, and Damon Sacco, MD

Abstract

The accuracy of using computed tomography (CT) to assess interbody fusion in patients with titanium implants has been questioned in the past. Radiologists have reported difficulty assessing fusion bone quality because of metal artifact and small graft windows. A new titanium interbody implant with a large footprint and a wide graft aperture has been developed.

We conducted a study to determine the interobserver reliability of using CT to assess radiographic fusion variables with the new titanium interbody device. Patients underwent anterior lumbar interbody fusion with the same titanium interbody implant. Reconstructed CT images were obtained randomly at 6, 9, or 12 months. Two independent radiologists reviewed the scans. Interobserver reliability was calculated using the κ statistic.

Fifty-six spinal fusion levels (33 patients) were analyzed. The radiologists agreed on 345 of the 392 fusion data points reviewed (κ = .88). Agreement for solid fusion formation was 0.77.

This interbody device demonstrated minimal artifact and minimal subsidence, and trabecular bone was easily identified throughout the implant in the vast majority of cases reviewed. High interobserver agreement was noted across all radiographic variables assessed.

The accuracy of using computed tomography (CT) to assess lumbar interbody fusion with titanium implants has been questioned in the past.[1-4] Reports have most often focused on older technologies using paired, threaded, smooth-surface titanium devices. Some authors have reported they could not confidently assess the quality of fusions using CT because of implant artifact.[1-3]

When pseudarthrosis is suspected clinically, and imaging results are inconclusive, surgical explorations may be performed with mechanical stressing of the segment to assess for motion.[2,5-7] However, surgical exploration not only has the morbidity of another surgery but may not be conclusive. Direct exploration of an interbody fusion is problematic. In some cases, there may be residual normal springing motion through posterior elements, even in the presence of a solid interbody fusion, which can be confusing.[5] Radiologic confirmation of fusion status is therefore preferred over surgical exploration. CT is the imaging modality used most often to assess spinal fusions.[8,9]

A new titanium interbody fusion implant (Endoskeleton TA; Titan Spine, Mequon, Wisconsin) preserves the endplate and has an acid-etched titanium surface for osseous integration and a wide central aperture for bone graft (**Figure 1**). Compared

Figure 1. Titanium interbody fusion implant (Endoskeleton TA; Titan Spine, Mequon, Wisconsin).

Authors' Disclosure Statement: This research was supported by a Spinal Research Foundation grant. Dr. Slosar wishes to report that he is Medical Director and a Scientific Advisory Board Member at Titan Spine, which makes the titanium interbody implant used in this study. The other authors report no actual or potential conflict of interest in relation to this article.

with earlier titanium implants, this design may allow for more accurate CT imaging and fusion assessment. We conducted a study to determine the interobserver reliability of using CT to evaluate bone formation and other radiographic variables with this new titanium interbody device.

Materials and Methods

After receiving institutional review board approval for this study, as well as patient consent, we obtained and analyzed CT scans of patients after they had undergone anterior lumbar interbody fusion (ALIF) at L3–S1 as part of a separate clinical outcomes study.

Each patient received an Endoskeleton TA implant. The fusion cage was packed with 2 sponges (3.0 mg per fusion level) of bone morphogenetic protein, or BMP (InFuse; Medtronic, Minneapolis, Minnesota). In addition, 1 to 3 cm^3 of hydroxyapatite/β-tricalcium phosphate (MasterGraft, Medtronic) collagen sponge was used as graft extender to fill any remaining gaps within the cage. Pedicle screw fixation was used in all cases.

Patients were randomly assigned to have fine-cut CT scans with reconstructed images at 6, 9, or 12 months. The scans were reviewed by 2 independent radiologists who were blinded to each other's interpretations and the clinical results. The radiographic fusion criteria are listed in **Tables 1 to 3**. Interobserver agreement (κ) was calculated separately for each radiographic criterion and could range from 0.00 (no agreement) to 1.00 (perfect agreement).[10,11]

Results

The study involved 33 patients (17 men, 16 women) with 56 lumbar spinal fusion levels. Mean age was 46 years (range, 23-66 years). Six patients (18%) were nicotine users. Seventeen patients were scanned at 6 months, 9 at 9 months, and 7 at 12 months. There were no significant differences in results between men and women, between nicotine users and nonusers, or among patients evaluated at 6, 9, or 12 months.

The radiologists agreed on 345 of the 392 data points reviewed (κ = 0.88). Interobserver agreement results for the fusion criteria are listed in **Tables 1 and 3**. Interobserver agreement was 0.77 for overall fusion grade, with the radiologists noting definite fusion (grade 5) in 80% and 91% of the levels (**Table 1**). Other radiographic criteria are listed in **Tables 2 and 3**. Interobserver agreement was 0.80 for degree of artifact, 0.95 for subsidence, 0.96 for both lucency and trabecular bone, 0.77 for anterior osseous bridging, and 0.95 for cystic vertebral changes.

Table 1. Fusion Grade

	Fusion Grade				
Reviewer	5: Definitely Fused	4: Probably Fused	3: Indeterminate	2: Probably Not Fused	1: Definitely Not Fused
A	45 (80%)	8 (14%)	2 (4%)	1 (2%)	0 (0%)
B	51 (91%)	5 (9%)	0 (0%)	0 (0%)	0 (0%)
Agreement					
Same Grade			0.77		
Within 1 Grade			0.95		

Table 2. Artifact Grades and Subsidence

	Artifact Grade			Subsidence	
Reviewer	0: Mild	1: Moderate	2: Nondiagnostic	Yes	No
A	43 (77%)	12 (21%)	1 (2%)	0 (0%)	56 (100%)
B	54 (96%)	2 (4%)	0 (0%)	3 (5%)	53 (95%)
Agreement		**0.80**		**0.95**	

Table 3. Other Fusion Assessment Criteria

	Lucency		Anterior Osseous Bridging		Trabecular Bone Throughout Cage		Cystic Changes in Vertebral Body	
Reviewer	Yes	No	Yes	No	Yes	No	Yes	No
A	1 (2%)	55 (98%)	27 (48%)	29 (52%)	54 (96%)	2 (4%)	2 (4%)	54 (96%)
B	2 (4%)	54 (96%)	39 (70%)	17 (30%)	56 (100%)	0 (0%)	3 (5%)	53 (95%)
Agreement	**0.96**		**0.77**		**0.96**		**0.95**	

Discussion

Radiographic analysis of interbody fusions is an important clinical issue. Investigators have shown that CT is the radiographic method of choice for assessing fusion.[8,9] Others have reported that assessing fusion with metallic interbody implants is more difficult compared with PEEK (polyether ether ketone) or allograft bone.[3,4,5,12]

Heithoff and colleagues[1,2] reported on difficulties they encountered in assessing interbody fusion with titanium implants, and their research has often been cited. The authors concluded that they could not accurately assess fusion in these cases because of artifact from the small apertures in the cages and metallic scatter. Their study was very small (8 patients, 12 surgical levels) and used paired BAK (Bagby and Kuslich) cages (Zimmer, Warsaw, Indiana).

Recently, a unique surface technology, used to manufacture osseointegrative dental implants, has been adapted for use in the spine.[13-15] Acid etching modifies the surface of titanium to create a nano-scale (micron-level) alteration. Com-

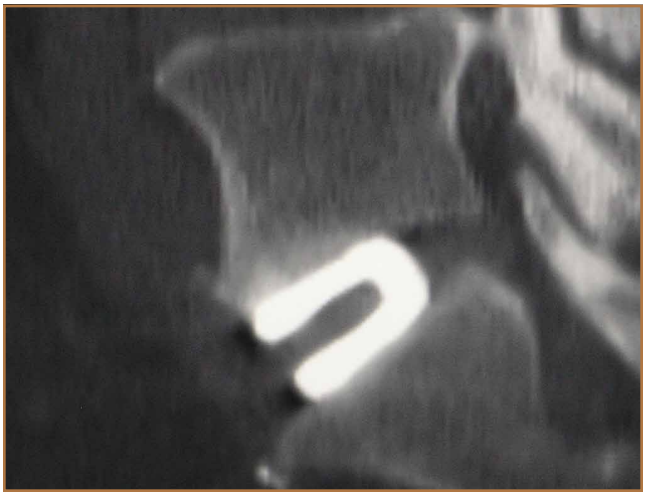

Figure 2. Example of grade 2 artifact (no subsidence or peri-implant lucencies).

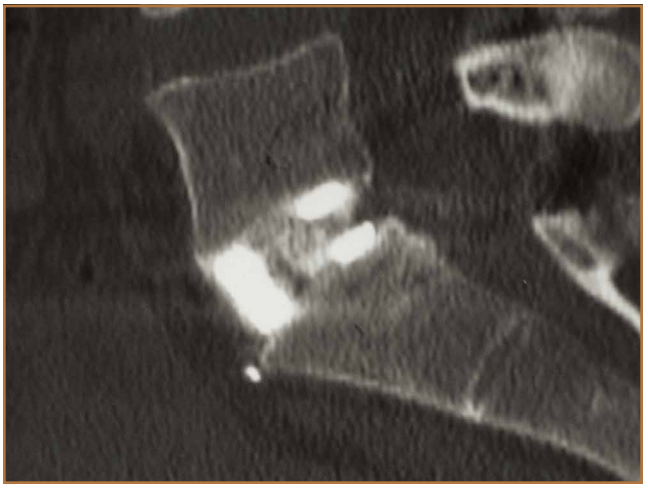

Figure 3. Example of grade 5 fusion.

Figure 4. Example of grade 4 fusion.

pared with PEEK and smooth titanium, acid-etched titanium stimulates a better osteogenic environment.[16,17] As this technology is now used clinically in spinal surgery, we thought it important to revisit the issue of CT analysis for fusion assessment with the newer titanium implants.

Artifact

The results of our study support the idea that the design of a titanium interbody fusion implant is important to radiographic analysis. The implant studied has a large open central aperture that appears to generate less artifact than historical controls (paired cylindrical cages) have.[1-4] Other investigators have reported fewer problems with artifact in their studies of implants incorporating larger openings for bone graft.[6,18] The radiologists in the present study found no significant problems with artifact. Less artifact is clinically important, as the remaining fusion variables can be more clearly visualized (**Table 2, Figure 2**).

Anterior Osseous Bridging, Subsidence, Lysis

In this study, the bony endplates were preserved. The disc and endplate cartilage was removed without reaming or drilling. Endplate reaming most likely contributes to subsidence and loss of original fixation between implant and bone interface.[1,4,12] Some authors have advocated recessing the cages deeply and then packing bone anteriorly to create a "sentinel fusion sign."[1,2,6] Deeply seating interbody implants, instead of resting them more widely on the apophyseal ring of the vertebral endplate, may also lead to subsidence.[4,12] The issue of identifying a sentinel fusion sign is relevant only if the surgeon tries to create one. In the present study, the implant used was an impacted cage positioned on the apophyseal perimeter of the disc space, just slightly recessed, so there was no attempt to create a sentinel fusion sign, as reflected in the relatively low scores on anterior osseous bridging (48%, 52%).

Subsidence and peri-implant lysis are pathologic variables associated with motion and bone loss. Sethi and colleagues[19] noted a high percentage of endplate resorption and subsidence in cases reviewed using PEEK or allograft spacers paired with BMP-2. Although BMP-2 was used in the present study, we found very low rates of subsidence (0%, 5%) and no significant peri-implant lucencies (2%, 4%) (**Figure 2**). Interobserver agreement for these variables was high (0.95, 0.96). We hypothesize that the combination of endplate-sparing surgical technique and implant–bone integration contributed to these results.

Trabecular Bone and Fusion Grade

The primary radiographic criterion for solid interbody fusion is trabecular bone throughout the cage, bridging the vertebral bodies. In our study, the success rates for this variable were 96% and 100%, and there was very high interobserver agreement (0.96) (**Figure 3**). This very high fusion rate may preclude detecting subtle differences in interobserver agreement, but to what degree, if any, is unknown. Other investigators have effectively identified trabecular bone across the interspace and

throughout the cages.[6,18] The openings for bone formation were larger in the implants they used than in first-generation fusion cages but not as large as the implant openings in the present study. Larger openings may correlate with improved ability to visualize bridging bone on CT.

Radiologists and surgeons must ultimately arrive at a conclusion regarding the likelihood a fusion has occurred. Our radiologists integrated all the separate radiologic variables cited here, as well as their overall impressions of the scans, to arrive at a final grade regarding fusion quality (**Figures 3, 4**). Although this category provides the most interpretive latitude of all the variables examined, the results demonstrate high interobserver reliability. Agreement to exactly the same fusion grade was 0.77, and agreement to within 1 category grade was 0.95.

This study had several limitations. Surgical explorations were not clinically indicated and were not performed. There were no suspected nonunions or hardware complications, two of the most common indications for exploration. In addition, this study was conducted not to determine specific accuracy of CT (compared with surgery exploration) for fusion assessment but to assess interobserver reliability. The clinical success rates for this population were high, and no patient required revision surgery for suspected pseudarthrosis. To assess the true accuracy of CT for fusion assessment, one would have to subject patients to follow-up exploratory surgery to test fusions mechanically.

Another limitation is the lack of a single industry-accepted radiographic fusion grading system. Fusion criteria are not standardized across all studies. Our radiologists have extensive research experience and limit their practices to neuromuscular radiology with a concentration on the spine. The radiographic criteria cited here are the same criteria they use in clinical practice, when reviewing CT scans for clinicians. Last, there was no control group for direct comparison against other cages. Historical controls were cited. This does not adversely affect the conclusions of this investigation.

Conclusion

Clinicians have been reluctant to rely on CT with titanium devices because of concerns about the accuracy of image interpretations. The interbody device used in this study demonstrated minimal artifact and minimal subsidence, and trabecular bone was easily identified throughout the implant in the majority of cases reviewed. We found high interobserver agreement scores across all fusion criteria. Although surgical exploration remains the gold standard for fusion assessment, surgeons should have confidence in using CT with this titanium implant.

Dr. Slosar is Attending Surgeon, SpineCare Medical Group; and Medical Director, San Francisco Spine Institute, Daly City, California. Dr. Kaiser is Attending Radiologist, National Orthopedic Imaging Associates, Greenbrae, California; and Assistant Clinical Professor, Department of Radiology, University of California, San Francisco, California. Dr. Marrero is Spine Fellow, SpineCare Medical Group, Daly City, CA. Dr. Sacco is Attending Radiologist, National Orthopedic Imaging Associates, Greenbrae, California.

Address correspondence to: Paul J. Slosar, MD, San Francisco Spine Institute, 455 Hickey Blvd., #310, Daly City, CA 94015 (tel, 650-985-7500; fax, 650-985-7506; e-mail, info@spinecare.com).

Am J Orthop. 2015;44(2):86-89.

References

1. Gilbert TJ, Heithoff KB, Mullin WJ. Radiographic assessment of cage-assisted interbody fusions in the lumbar spine. *Semin Spine Surg.* 2001;13:311-315.
2. Heithoff KB, Mullin WJ, Renfrew DL, Gilbert TJ. The failure of radiographic detection of pseudarthrosis in patients with titanium lumbar interbody fusion cages. In: Proceedings of the 14th Annual Meeting of the North American Spine Society; October 20-23, 1999; Chicago, IL. Abstract 14.
3. Cizek GR, Boyd LM. Imaging pitfalls of interbody implants. *Spine*. 2000;25(20):2633-2636.
4. Dorchak JD, Burkus JK, Foor BD, Sanders DL. Dual paired proximity and combined BAK/proximity interbody fusion cages: radiographic results. In: *Proceedings of the 15th Annual Meeting of the North American Spine Society.* New Orleans, LA: North American Spine Society; 2000:83-85.
5. Santos ER, Goss DG, Morcom RK, Fraser RD. Radiologic assessment of interbody fusion using carbon fiber cages. *Spine.* 2003;28(10):997-1001.
6. Carreon LY, Glassman SD, Schwender JD, Subach BR, Gornet MF, Ohno S. Reliability and accuracy of fine-cut computed tomography scans to determine the status of anterior interbody fusions with metallic cages. *Spine J*. 2008;8(6):998-1002.
7. Fogel GR, Toohey JS, Neidre A, Brantigan JW. Fusion assessment of posterior lumbar interbody fusion using radiolucent cages: x-ray films and helical computed tomography scans compared with surgical exploration of fusion. *Spine J*. 2008;8(4):570-577.
8. Selby MD, Clark SR, Hall DJ, Freeman BJ. Radiologic assessment of spinal fusion. *J Am Acad Orthop Surg.* 2012;20(11):694-703.
9. Chafetz N, Cann CE, Morris JM, Steinbach LS, Goldberg HI, Ax L. Pseudarthrosis following lumbar fusion: detection by direct coronal CT scanning. *Radiology.* 1987;162(3):803-805.
10. Landis RJ, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
11. Viera AJ, Garrett JM. Understanding interobserver agreement; the kappa statistic. *Fam Med.* 2005;37(5):360-363.
12. Burkus JK, Foley K, Haid RW, Lehuec JC. Surgical Interbody Research Group—radiographic assessment of interbody fusion devices: fusion criteria for anterior lumbar interbody surgery. *Neurosurg Focus*. 2001;10(4):E11.
13. Albrektsson T, Zarb G, Worthington P, Eriksson AR. The long-term efficacy of currently used dental implants: a review and proposed criteria of success. *Int J Oral Maxillofac Implants.* 1986;1(1):11-25.
14. De Leonardis D, Garg AK, Pecora GE. Osseointegration of rough acid-etched titanium implants: 5-year follow-up of 100 Minimatic implants. *Int J Oral Maxillofac Implants*. 1999;14(3):384-391.
15. Schwartz Z, Raz P, Zhao G, et al. Effect of micrometer-scale roughness on the surface of Ti6Al4V pedicle screws in vitro and in vivo. *J Bone Joint Surg Am*. 2008;90(11):2485-2498.
16. Olivares-Navarrete R, Gittens RA, Schneider JM, et al. Osteoblasts exhibit a more differentiated phenotype and increased bone morphogenetic protein production on titanium alloy substrates than on poly-ether-ether-ketone. *Spine J*. 2012;12(3):265-272.
17. Olivares-Navarrete R, Hyzy SL, Gittens RA 1st, et al. Rough titanium alloys regulate osteoblast production of angiogenic factors. *Spine J*. 2013;13(11):1563-1570.
18. Burkus JK, Dorchak JD, Sanders DL. Radiographic assessment of interbody fusion using recombinant human bone morphogenetic protein type 2. *Spine*. 2003;28(4):372-377.
19. Sethi A, Craig J, Bartol S, et al. Radiographic and CT evaluation of recombinant human bone morphogenetic protein-2–assisted spinal interbody fusion. *AJR Am J Roentgenol.* 2011;197(1):W128-W133.