The maturation of grafted bone after posterior lumbar interbody fusion with an interbody carbon cage

A PROSPECTIVE FIVE-YEAR STUDY

We evaluated the maturation of grafted bone in cases of successful fusion after a one- or two-level posterior lumbar interbody fusion (PLIF) using interbody carbon cages. We carried out a five-year prospective longitudinal radiological evaluation of patients using plain radiographs and CT scans. One year after surgery, 117 patients with an early successful fusion were selected for inclusion in the study. Radiological evaluation of interbody bone fusion was graded on a 4-point scale. The mean grades of all radiological and CT assessments increased in the five years after surgery, and differences compared to the previous time interval were statistically significant for three or four years after surgery. Because the grafted bone continues to mature for three years after surgery, the success of a fusion should not be assessed until at least three years have elapsed. There were no significant differences in the longitudinal patterns of grafted bone maturity between iliac bone and local bone. However, iliac bone grafting may remodel faster than local bone.

Patients and Methods

Between April 2003 and December 2005, 154 consecutive patients underwent PLIF with pedicle screws and interbody carbon cages for one- or two-level degenerative disease of the lumbar spine. Ethical committee approval for the study had previously been obtained. All patients met the inclusion criteria of symptomatic one- or two-level degenerative disease (L2/L3 to L5/S1) requiring one or two-level PLIF using pedicle screws and carbon cages with autogenous local bone with or without additional iliac bone graft. Exclusion criteria included a PLIF at more than two levels, a lumbar kyphosis (lordosis L1/S < 0°) or spondylolisthesis > grade III (> 50%), a previous attempt at fusion at the intended surgical level, a medical condition that required corticosteroids, severe osteoporosis, autoimmune disease, malignancy or infection. In total, 144 patients met these criteria (Fig. 1). We selected cases of early successful arthrodesis at 12 months after surgery using lateral flexion-extension radiographs and sagittal and coronal reconstructed thin-section helical CT scans. An early successful fusion was defined...
as showing continuous bony bridging (but not necessarily trabecular bone) between the upper and lower vertebrae, an absence of > 5° angular movement on lateral flexion-extension radiographs and an absence of radiolucency around the pedicle screws or cages. Based on these criteria, 128 patients were selected for follow-up (Fig. 1). The maturation of their grafted fusion bone was followed radiologically for five years. Sixteen patients were excluded because of early pseudarthrosis and will be addressed in another study.

Surgical procedures. The same surgical technique was used in each case by one of three surgeons (TK, GY and YS). A complete posterior neural decompression, including laminectomy, total or partial facetectomy and discectomy, was performed. Interbody fusion was carried out using two carbon fibre cages (DePuy AcroMed Corp., Raynham, Massachusetts) filled with local bone obtained during decompression with or without the addition of iliac crest bone. The bone graft was tightly packed around and between the cages. An image intensifier was used to place the carbon cages in the disc space close to the anterior cortical margin of the vertebrae. The fusion was then stabilised with four pedicle screws and rods with the application of an appropriate compressive force. Patients were mobilised on the second post-operative day in a low-profile thoracolumbosacral orthosis (TLSO), which was retained for three months.

Radiological assessment and grading. Anteroposterior (AP) and lateral radiographs, which included flexion-extension views, and thin-section helical CT scans were obtained at one, three, six, 12 and 18 months and yearly thereafter until five years after surgery. Fusion was assessed radiologically by two independent spinal surgeons (YI and AM), with a third reviewer (TK) available for adjudication. AP radiographs were taken by the Ferguson method19: the X-ray beam was directed parallel to the end-plates of the fused level. A 1-mm thin-section helical CT scan of the involved lumbar segments was obtained. Reconstructed sagittal and coronal CT scans were generated perpendicular to the plane of each cage.

The radiological state of the fusion was assessed according to the difference between the radiological density of the bone and four square carbon cage struts on the AP radiographs (cross sign), the presence or absence of continuous bony bridging on lateral radiographs, the extension of bony bridging anterior and posterior to the cage on lateral radiographs, continuous bridging on sagittal CT scans, the extension of bony bridging fusion density on sagittal CT scans and the status of remodelling of the grafted bone to the trabecular bone on CT scans. Radiological and CT findings were graded on a 4-point scale (0 to 3; Figs 2 and 3).

Statistical analysis. Statistical analysis was performed using SPSS 17.0 (SPSS Inc., Tokyo, Japan). Before the study began, the radiological scoring method had been carried out twice with an interval of three weeks in 20
randomly selected patients to evaluate intra- and interobserver reliability. We used kappa statistics to describe intra- and interobserver agreement using the method described by Landis and Koch.\textsuperscript{24,25}

We divided the patients into two groups for the purposes of analysis: those in whom local bone with iliac crest bone (ICB) was used and those in whom local bone without iliac crest bone (LB) was used. A comparative analysis of the status of the grafted bone between the two groups was performed by two-way repeated-measures analysis of variance (ANOVA) by ranks. Changes in the status of the grafted bone compared to the previous interval for each radiological assessment within a group were determined using the Mann-Whitney U test. Categorical data were analysed using the chi-squared and Fisher’s exact tests, as appropriate. The Mann-Whitney U test was applied for all variables with ordinal categorisation, and p < 0.05 was considered statistically significant.

Results

Of the 128 patients with early successful fusion, 117 were available for five-year follow-up. Complete radiological and clinical data at each interval over the five years were obtained for 82 patients, and for 30 patients > 70% of the data was obtained. The remaining five patients for whom < 70% of the data was available were excluded. A total of 112 (72.7%) patients were finally included in the longitudinal radiological assessment (Fig. 1 and Table I).

Intra- and interobserver reliability. For intra-observer agreement for the cross sign on AP radiographs the kappa value was 0.72; for lateral bony bridging and anterior and posterior extension of bony bridging on lateral radiographs the kappa values were 0.62, 0.59, and 0.58, respectively, and for sagittal bony bridging, anterior and posterior extension of bony bridging, and remodelling status of the grafted bone on CT scans they were 0.92, 0.84, 0.77 and 0.72, respectively.

For interobserver agreement for the cross sign on AP radiographs the kappa value was 0.45; for lateral bony bridging and anterior and posterior extension the kappa values were 0.45, 0.65 and 0.45, respectively, and for sagittal bony bridging, anterior and posterior extension, and remodelling status on CT scans they were 0.47, 0.50, 0.59 and 0.41, respectively.

Fusion bone maturity in cases of successful arthrodesis. An overall comparative analysis showed no significant differences between the ICB and LB groups in the longitudinal pattern of mature graft bone (Figs 4 to 10).

The mean grades in both groups for all radiological assessments tended to increase at each time interval compared to the previous interval, with the exception of the cross sign on AP radiographs. The mean grade for the cross sign in each group increased significantly compared to the previous time interval, up to one year (p < 0.01), but then decreased significantly up to five years in the ICB group (two and four years, p < 0.01; three years, p = 0.048; five years, p = 0.018) and three years in the LB group (three
years, \( p = 0.044 \); Fig. 4). The mean grade for continuous bony bridging on lateral radiographs increased significantly up to three years (ICB group: three months, \( p = 0.049 \); six months, one, two and three years, \( p < 0.01 \); LB group: six months, \( p = 0.011 \); three years, \( p = 0.048 \); one and two years, \( p < 0.01 \); Fig. 5). However, CT scans revealed that the mean grade for continuous bony bridging tended to decrease at three months, and then increase significantly from six months to three years (ICB group: one, two and three years, \( p < 0.01 \); LB group: one year, \( p = 0.027 \); three years, \( p = 0.049 \); two years, \( p < 0.01 \); Fig. 7). The mean grade for the remodelling status of the grafted bone to the trabecular bone on CT scans.

| Table I. Demographic data (ICB, iliac crest bone; LB, local bone) |
|---------------------------------|-----------------|-----------------|-----------------|
|                                | Total           | ICB group       | LB group        |
| No. of patients                | 112             | 63              | 49              |
| Mean age (yrs) (range)         | 59.2 (31 to 70) | 58.6 (31 to 70) | 60.1 (33 to 70) |
| Gender (% male)                | 55.4            | 50.8            | 61.2            |
| No. of fusion level (1:2 level) | 98:14           | 56.7            | 42.7            |
| Fusion level                   |                 |                 |                 |
| L2/3                            | 1               | 1               | 0               |
| L3/4                            | 22              | 10              | 12              |
| L4/5                            | 88              | 52              | 36              |
| L5/S                            | 15              | 7               | 8               |
| Mean follow-up (months) (range)| 75.8 (62 to 91) | 75.5 (62 to 91) | 76.5 (62 to 91) |
| p-value                         |                 |                 | 0.6283          |

Interpretations of CT scans were graded on a 4-point scale (0 to 3). (a) Continuous bony bridging between the end-plates of the upper and lower vertebrae and grafted bone on sagittal CT scans. (b) Extension of bony bridging anterior and posterior to the cage on sagittal CT scans. (c) Remodelling status of the grafted bone to the trabecular bone on CT scans.
trabecular bone on CT scan to four years in the ICB group and to three years in the LB group were significant (p < 0.01; Fig. 10).

Significant remodelling was seen up to four years (p < 0.01) in the ICB group and three years (three months to two years, p < 0.01; three years, p = 0.046) in the LB group. One year after surgery, 39% of the 70 levels in the ICB group and 8.6% of the 56 levels in the LB group showed >50% of trabecular bone formation in the original bone graft area (points 2 and 3); two years after surgery, 70% in the ICB group and 51.5% in the LB group; and three years after surgery, 90% in the ICB group and 77.1% in the LB group. Five years after surgery only 5% of levels showed < 50% trabecular bone formation in the original bone graft area in the ICB group; however, in the LB group these remained at 22.9% (p = 0.026).

Discussion
Several studies of PLIF using a variety of pedicle screws and interbody cages have reported different rates of fusion at different post-operative time intervals.2,3,5,7,14-19,22 A 2001 symposium revealed considerable disagreement among eight different authors about the best method to use, and
the criteria with which to assess fusion. In the intervening ten years no universally accepted radiological assessment criteria have been reported.

The assessment of fusion after PLIF with metal cages has been dependent on lateral flexion–extension radiographs. The absence of movement does not, however, necessarily correspond with solid fusion. Sagittal and coronal reconstructed helical CT scans are better at assessing interbody fusion when a metal cage has been implanted. However, it has been argued that the bone inside the cage cannot be accurately assessed either by plain radiographs or by CT scans.

Radiolucent interbody cages allow bony bridging to be assessed radiologically. Brantigan et al. showed that plain radiographs were both highly sensitive and predicted fusion accurately when carbon cages had been used. As a result of this, Diedrich et al. reported a moderate disparity in intervertebral fusion rates between radio-opaque and radiolucent cages. Fogel et al. concluded that helical CT is unlikely to provide useful further information when plain radiographs show strong evidence of fusion or pseudarthrosis.

In our study, all kappa values for intra- and interobserver agreement for plain radiographs and CT scans were classified as moderate, substantial or perfect. These values were acceptable, as we expected extreme variability in the interpretation of both plain radiographs and CT scans. The classification method we used has been adopted by many spinal surgeons worldwide. The use of radiolucent material for interbody cages may therefore improve inter- and intra-study correlations of radiological fusion after PLIF. Furthermore, although the kappa values in this study were acceptable, those for interobserver agreement were slightly low. Therefore, by improving interobserver agreement on the interpretation of plain radiographs and CT scans, it should be possible to refine our assessment of a successful fusion.

The mean grades for all measurements up to five years after surgery showed a tendency to increase, and the differences between these grades at each time interval compared to the previous interval were statistically significant until three or four years after surgery. Furthermore, only 51.5% of the LB group and 70% of the ICB group showed > 50% fusion.

ANOVA p = 0.270

**Fig. 8**

Longitudinal changes in extension of bony bridging fusion density anterior to the cage on CT scans. Mean anterior extension grades increased significantly until three years. *1 p = 0.027, *2 p = 0.042, *3 p = 0.026, and ** p < 0.01, Mann-Whitney U test. ANOVA, analysis of variance.

ANOVA p = 0.141

**Fig. 10**

Longitudinal changes in the remodelling status of grafted bone to trabecular bone on CT scans. Mean grades increased until five years, and increases from three months to four years in the ICB group and to three 3 years in the LB group were significant. *1 p < 0.01, Mann-Whitney U test. ANOVA, analysis of variance.
trabecular bone formation in the original bone graft area two years after surgery. Nevertheless, the proportion of grades that showed > 50% trabecular bone formation in the original bone graft area increased significantly to 77.1% in the LB group and to 90% in the ICB group three years after surgery. Diedrich et al. reported similar results based on their criteria for radiological fusion. According to them, it is doubtful whether it is possible to determine the success of fusion or the presence of a pseudarthrosis within two years of surgery. A final decision on the success of fusion after a PLIF using interbody cages should therefore be made at least three years after surgery.

Numerous researchers have noted that CT is better than plain radiographs for assessing fusion, whether metallic or non-metallic interbody cages are used. Although the consensus is that thin-section helical CT is the best technique for identifying bony bridging, there is concern that it may overestimate its presence if carried out early after surgery. In this study, the mean grades on CT interpretation were > 1 point in the early post-operative period, but with no significant increase for the next six months. When much of the grafted bone is tightly packed into and around the cage, sagittal or coronal thin-section CT scans show fine continuous bony bridging without trabeculation during the early post-operative period. If thin-section CT is not used, a small radiolucent area in which the grafted bone was not packed during surgery may not be apparent. These results show that the interpretation of continuous bony bridging with CT may lead to a false assessment of fusion status within six months of surgery.

Some researchers have used bone obtained during posterior decompression as graft, and have reported fusion rates similar to those achieved using autologous iliac bone graft. We have already reported that almost the same fusion rates can be obtained using local bone graft and autologous iliac bone graft in patients who undergo a PLIF using carbon cages. However, whether local bone is an alternative to iliac bone grafting is still controversial because there are no prospective randomised control studies comparing fusion rates using long-term longitudinal radiological evaluation. This is the first study to compare the long-term graft maturity achieved by local and iliac bone grafting in patients with a successful fusion. Our overall comparative analysis shows no significant difference between the two groups. However, significant differences were found between the two groups, in both level and grade, in the remodelling status of the grafted bone to the trabecular bone on CT scans. After one year, the levels showing > 50% trabecular bone formation in the original bone graft area (points 2 and 3) were 39% in the ICB group and 8.6% in the LB group (p < 0.001). After two years, more than two-thirds of the levels in the ICB group showed > 50% trabecular bone formation, whereas in the LB group half of the levels showed > 50% trabecular bone formation. After five years, only 5% of the ICB group showed < 50% trabecular bone formation in the original bone graft area, but in the LB group it remained at 22.9% (p = 0.026). Therefore, although the longitudinal patterns of the maturity of the grafted bone were not significantly different between the two methods, the use of iliac bone may result in remodelling to trabecular bone faster than that achieved by local bone.

The main limitation of this study is that it only evaluates the long-term maturity of grafted bone in cases of successful fusion, and excludes those with an early pseudarthrosis. It was not designed to compare the fusion rates between the two bone grafting methods. A prospective randomised control trial would be needed to show that autologous iliac bone graft is better than local bone. Furthermore, because 16 patients had an early pseudarthrosis and because we considered this number too small for a comparative study, a further study is needed to determine the results in this group. Finally, the aim of this study was to evaluate the maturity of grafted bone in patients with a successful PLIF using interbody carbon cages, and not to assess the clinical outcome of PLIF. A successful PLIF does not necessarily correspond to a successful outcome for the patient. A successful outcome in patients after PLIF is the result of judicious patient selection and meticulous surgery, and not just a successful fusion.

**Supplementary material**

Tables showing i) the grading system of interbody bone fusion sites of radiographs and CT scans, ii) the kappa values for inter- and intra-observer agreement for each radiological assessment, and iii) the proportion of grading of levels in the remodelling status of the grafted bone to the trabecular bone on CT scans are available with the electronic version of this article on our website www.jbjs.org.uk

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

**References**