

CLINICAL ARTICLE

Relationship of Surgical Accuracy and Clinical Outcomes in Charitè Lumbar Disc Replacement

Christopher W Jones MBBS(Hons), PhD^{1,2}, Peter Smitham MBBS(Hons)², William R Walsh PhD²

¹Royal Prince Alfred Hospital, University of Sydney, and ²Surgical & Orthopaedic Research Laboratory, Prince of Wales Hospital, Sydney, New South Wales, Australia

Objective: To retrospectively assess the clinical and functional outcomes of a group of patients receiving Charitè lumbar disc replacement and to compare those outcomes to the corresponding surgical technical accuracy.

Methods: A retrospective study of all patients treated over a 3-year period was undertaken. Objective pain scores were quantified from 1 to 10. Short Form 36-Health Survey (SF-36v2) scores were compared to Australian population norms. Surgical placements were radiographically classified. Heterotopic ossification, disc height restoration and angle-defined instability were assessed using established protocols.

Results: Twenty-five patients were identified with three patients lost to follow-up. Average follow-up was 34 months. Ideal surgical placement was achieved in five (33%) single-level and three (37.5%) dual-level disc replacements. Sub-optimal surgical placement was seen in nine (60%) single-level and five (62.5%) dual-level disc replacements. Poor surgical placement was observed in a single-level disc replacement. All patients demonstrated a reduction in objective pain score ($P < 0.05$). SF-36v2 outcomes were superior in single-level compared to dual-level and ideal compared to sub-optimal replacements ($P < 0.05$).

Conclusion: The hypothesis that ideal surgical placements are associated with improved clinical and functional outcomes in total lumbar disc replacement was confirmed.

Key words: Arthroplasty; Intervertebral disk; Lumbar vertebrae; Pain; Replacement; Spine

Introduction

Lumbar disc replacement has been indicated as an alternative to lumbar fusion for patients suffering from intractable lower back pain who are unresponsive to non-operative treatment. While arthrodesis remains the major surgical approach to managing degenerative disc disease, the technique has been shown to alter the stress distribution in adjacent segments, possibly accelerating degenerative processes therein¹. The dual aims of disc replacement are therefore to provide adequate palliation of long-term back pain and to restore functional disc motion, thereby preserving adjacent vertebrae. Anterior approach lumbar disc replacement addresses several limitations of posterior approach lumbar

fusion by preserving paraspinous soft tissues, avoiding graft harvest site morbidity and reducing the incidence of adjacent segment disease via the restoration of near physiologic joint motion². Currently there are four principal total lumbar disc devices commercially available; the SB Charitè III (Depuy Spine, Johnson and Johnson, Raynham, MA, USA), the ProDisc II (Spine Solutions, New York, NY, USA), the Maverick (Medtronic Sofamor Danek, Memphis, TN, USA) and the Flexicore (SpineCore, Summit, NJ, USA). Of these devices, the Charitè has received Food and Drug Administration (FDA) approval following investigational device exemption (IDE) randomized-control trials, and has been the most used, particularly in Europe^{1,2}.

Address for correspondence William R Walsh, PhD, BEng (Hons), Surgical & Orthopaedic Research Laboratory, Prince of Wales Hospital, University of New South Wales, Randwick, Sydney, NSW 2031, Australia Tel: 0061-2-9382-2657; Fax: 0061-2-9382-2660; Email: w.walsh@unsw.edu.au

Disclosure: This project was independently funded by the Surgical & Orthopaedic Research Laboratory of UNSW and received no additional sources of support in the form of grants, equipment, and/or pharmaceutical items.

Received 2 March 2012; accepted 27 June 2012

TABLE 1 Patient selection criteria for lumbar disc replacement surgery

- Mechanical back pain of greater than 1-year duration remaining refractory to other treatments
- Mechanical back pain with neurological involvement (sciatic or bladder/bowel dysfunction)
- Radiographic or magnetic resonance imaging (MRI) demonstrated disc degeneration (endplate changes on T1 & T2 weighted MRI and/or desiccated disc on T2 weighted MRI)
- Facet joint involvement excluded by single photon emission computed tomography scan (SPECT scan)
- Age 18–60 years
- Suitable for anterior surgical approach
- No contraindicated co-morbidities (i.e. neoplastic disease)
- Willing and able to give informed consent

The Charitè is an artificial disc composed of a polyethylene core between cobalt-chromium-molybdenum plasma-sprayed porous coated endplates. Fixation is achieved via interference fit between the rostral and caudal vertebral body endplates, with the aid of six locator “fins”³. The Charitè disc design enables a range of motion (ROM) in six degrees of freedom and provides a small amount of axial compressibility¹. The results of lumbar disc replacement with the Charitè prosthesis have been examined by a number of previous studies, with excellent radiographic and functional outcomes reported^{3–7}. However, reports remain limited regarding the relationship between radiological appearance, surgical technical accuracy and clinical outcomes^{3,6}.

We hypothesized that ideal surgical placements are associated with improved outcomes in total lumbar disc replacement; therefore the aim of this study was to retrospectively assess the clinical and functional outcomes of a group of patients receiving Charitè lumbar disc replacement and to compare those outcomes to the corresponding surgical technical accuracy. The influence of a number of factors upon the degree of surgical technical accuracy achieved was also examined.

Materials and Methods

Study Design

A retrospective review of patients treated over a 3-year period with Charitè single-level or dual-level lumbar disc replacement was undertaken. A single orthopaedic surgeon in Australia operated on all patients recruited into this study. Surgical technique was consistent between all patients and executed according to manufacturer and FDA guidelines.

Patient Selection Criteria

Prior to being considered for lumbar disc replacement surgery, patients were required to satisfy the following criteria (Table 1).

Surgical Technique

The surgical technique described by Geisler (2005) was used in all operations⁸. Briefly, the patient was placed supine on a folding operating table to enable intra-operative segmental lordosis adjustment. During the approach, a vascular surgeon was used to facilitate spinal access and reduce vascular com-

plications. Fluoroscopy was used to identify the approach angle and the location of the disc space, with access achieved via a standard left-sided retroperitoneal approach. Vascular retraction enabled the relevant disc space to be exposed; mobilization of the aorta, vena cava, sympathetic chain and left iliac artery and vein enabled L4/L5 exposure, while L5/S1 exposure was achieved via mobilization and retraction of the left and right iliac arteries and veins. Complete discectomy was then performed using standard instrumentation, enabling distraction and implant site preparation with curettage of the cartilaginous vertebral endplates. The approach described above enabled the artificial disc to be inserted along a midline trajectory using the Centreline TDR instruments (DePuy Spine), directly in the central axis of the disc space. Positioning was confirmed intraoperatively, and then using live fluoroscopy the disc endplates were inserted. Disc space distraction/height restoration was achieved using spreading and insertion forceps. Core trials were tested before final disc core selection, with final prosthesis positioning confirmed fluoroscopically.

Follow-up Methodology

We attempted to follow all patients who had received lumbar disc replacement with the Charitè prosthesis at our centre over a 3-year period. Data were collated in three stages over 6 months. Initially patients were mailed an information package containing a cover letter explaining the aims of the study, a postage-paid preaddressed return envelope, and an objective pain score survey and the Short Form 36-Version 2 Health Survey (SF-36v2) questionnaire. Patients were then contacted by phone, with reminders to complete the questionnaire and arrangements made for radiological and clinical assessments. Finally, all patients who had not responded by mail and/or had not attended clinical assessments were contacted by phone and interviewed using the SF-36v2 assessment tool. These patients were also sent radiological referrals for their local area with arrangements made for radiographs to be forwarded to the study centre for scanning, grading and interpretation.

Radiographic Scanning and Analysis

Static anteroposterior (AP) and lateral (Lat) radiographs were obtained and compared to preoperative or intraoperative films. Where possible, dynamic flexion and extension radiographs were also obtained. In accordance with established techniques, all radiographs were scanned on a UMAX PowerLook

2100XL (Techville, Dallas, TX, USA) flatbed scanner with images acquired through the UMAX proprietary software MagicScan 32 V4.5 (Techville, Dallas, TX, USA)^{6,9}. Scans were acquired in transmissive mode, at 400 DPI in 256 greyscale, with no magnification or filter and output in standard TIF digital format. Images were processed in Adobe Photoshop CS (Adobe Systems, San Jose, CA, USA) and the greyscale levels adjusted to maximize bone to soft tissue contrast. Following conversion to RGB format, working lines (Fig. 1) were drawn to establish the (caudal and rostral) joint surface (blue) and the position of the vertebral body centroid (green). The central axis (red) was then determined by a working line through the centroid intersection and perpendicular to the joint surface. Surgical placement accuracy and restoration of disc height were both then determined relative to this central axis.

Using the technique of McAfee *et al.*, the accuracy of surgical placement of the Charité artificial disc was classified as Group I ideal, Group II sub-optimal or Group III poor according to the anteroposterior and lateral radiographic appearance. In order to overcome any errors related to radiographic magnification variation, absolute measurements were not calculated; rather the position of the Charité artificial disc relative to the exact central axis was considered, with the central position determined by the position of the Charité locator “fins” in the AP view and the disc centroid in the lateral Lat view. Group I ideal was defined as the Charité artificial disc central locator fin (AP) or disc centroid (Lat) placed exactly on the central axis (Fig. 1A,B). Group II sub-optimal was defined as the central axis passing between lateral and central locator fins (AP) and the central axis passing through the disc at a point not more than a 1/2 radius from the centroid of the disc (Lat) (Fig. 1C,D). Group III poor was defined as the central axis (AP) passing lateral to the lateral locator fins and the central axis (Lat) passing through the disc at a point more than 1/2 radius from the centroid of the disc (Fig. 1E,F).

Heterotopic ossification was radiographically assessed on the basis of AP and Lat radiographs and classified as: no formation, marginal osteophytes, unilateral bridging, bilateral bridging or circumferential formation (Fig. 2). National Health and Medical Research Council Australia (NHRMC) definitions of weight relative to body mass index (BMI) were used: underweight = BMI < 20; healthy weight = BMI ≥ 20 and ≤ 25; overweight = BMI ≥ 25 and ≤ 30; obese = BMI > 30.

As established by Mochida *et al.*, relative vertebral body (A) and intervertebral disc (a) heights were measured using the Adobe Photoshop CS pixel measuring tool (Fig. 3)¹⁰. As described above, the advantage of this relative measuring technique was that no correction for radiographic magnification differences was required. In instances where preoperative (or intraoperative) radiographs were available, disc height restoration coefficient (DHR) was calculated according to Equation 1.

$$DHR = \frac{d^a}{\text{Pre} \frac{a}{A}} \quad (1)$$

The angle of intervertebral body instability was measured using the Adobe Photoshop CS angle measurement tool via the intersection of the working lines tracing the joint surfaces in the anteroposterior view as defined in Fig. 3¹⁰.

Clinical Outcome Measures

The standardized Short Form 36-Version 2 Health Survey (SF-36v2) (Medical Outcomes Trust, Quality Metric, Lincoln, RI, USA) was used to assess physical, mental and overall health of patients following lumbar disc replacement. The SF-36 Health Survey is currently recognized as the most widely used health status survey in the world and Australian norms have been recently published^{11,12}. Where possible this survey was completed by participants and returned by mail; however, in incidences where no reply was forthcoming, phone interviews were then conducted as detailed above. Objective pain scores were quantified using a 0–10 intensity scale (0 = no pain, 6 = severe pain, 10 = worst possible pain), with patients asked to retrospectively score their pain both immediately prior to their disc replacement and currently (at the time of responding); the numerical difference between these scores was defined as the Objective Pain Score Reduction. Additional quality of life variables such as a return to employment were also reported as practical measures.

Statistical Analysis

All statistical analysis was performed using SPSS V11.0.4 (SPSS, Chicago, IL, USA), using independent samples *t*-tests to examine differences in the dependent variable for independently grouped data. Levene’s test for equality of variances was used as a standard pre-test measure of methodological suitability. For the primary outcomes of objective pain score reduction and SF-36v2 score, data were grouped according to surgical technical accuracy, level of disc replacement, heterotopic ossification, BMI, disc height restoration and angle-defined stability. Differences in the primary outcomes of objective pain score reduction and SF-36v2 score were then tested for each independent grouping. SF-36v2 scores for each surgical technical accuracy classification were compared to Australian population norms from the 2004 South Australian Health Omnibus Survey (SAHOS)¹². Similarly for the secondary outcome of surgical technical accuracy, data were grouped according to level of disc replacement, heterotopic ossification, BMI, disc height restoration and angle-defined stability. Single-level and dual-level lumbar disc replacements were generally considered separately. However, for comparing differences in surgical technical accuracy according to the groupings of heterotopic ossification, BMI and angle defined stability, the analysis of single-level and dual-level replacements was combined by considering the surgical accuracy of each dual-level placement singularly, thereby increasing the sample size from $n = 19$ to $n = 23$ in an effort to improve power.

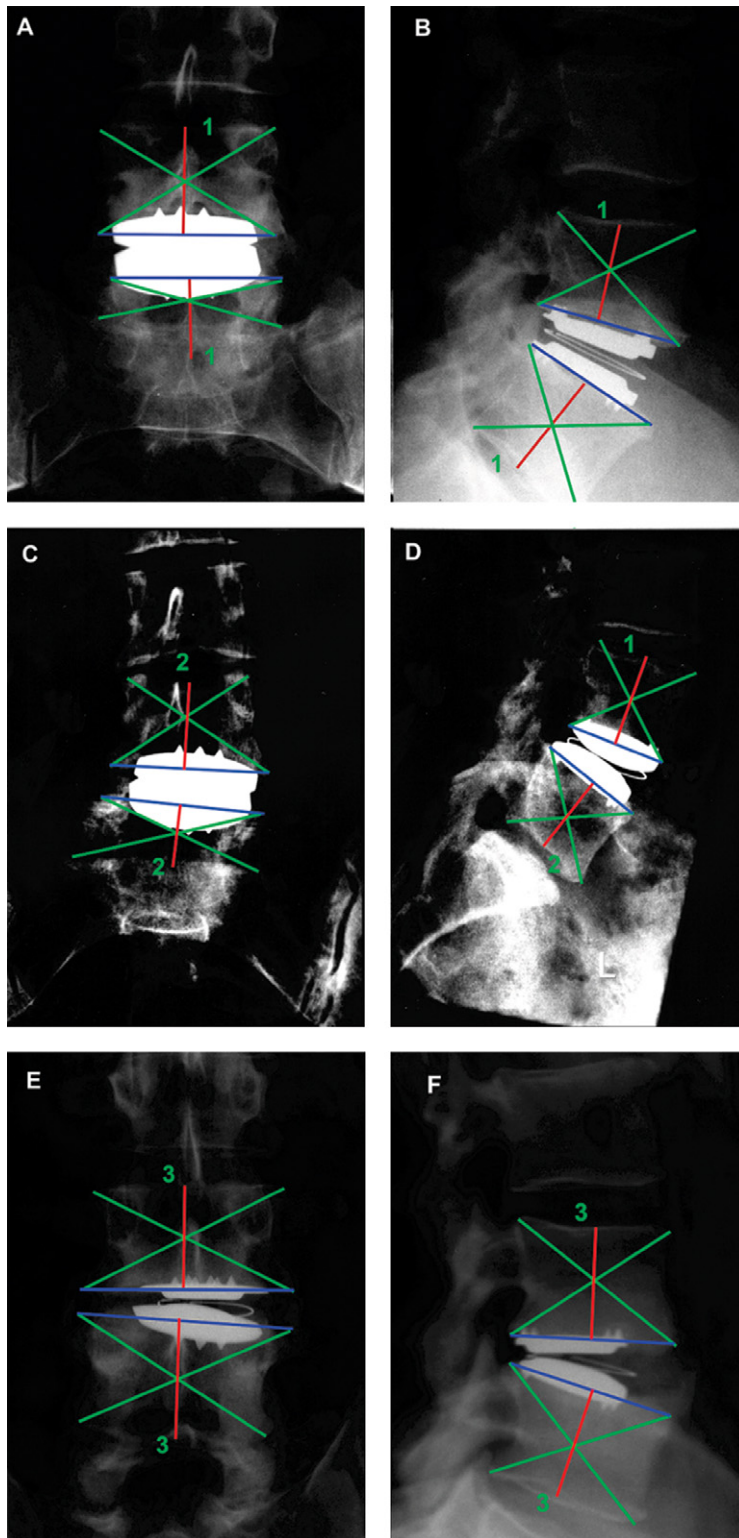


Fig. 1 Radiographic classification of surgical technical accuracy of L4/L5 Charitè disc replacement. Ideal placement was shown in both antero-posterior (A) and lateral (B) views, with prosthesis located at centroid of both caudal and rostral vertebral bodies. Sub-optimal placement was demonstrated in both antero-posterior (C) and lateral (D) views, with the prosthesis centroid offset laterally, but still within manufacturers' guidelines. Poor surgical placement was shown, with prosthesis offset beyond lateral disc "tooth" in antero-posterior view (E) and posterior to the centroid in the lateral view (F).

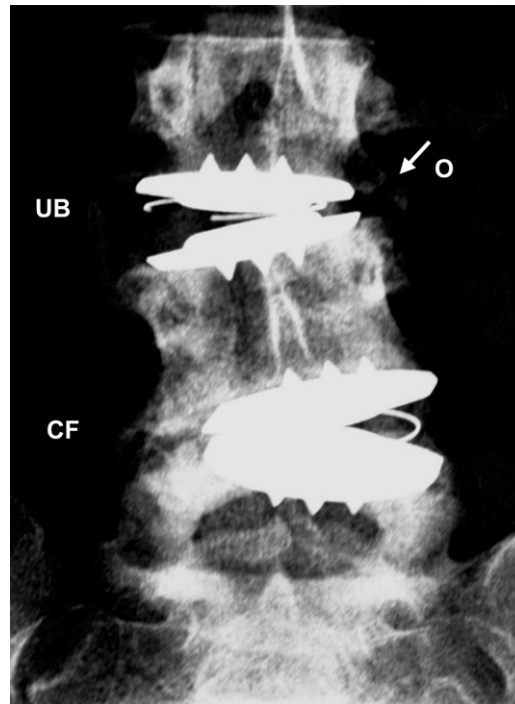


Fig. 2 Assessment of heterotopic ossification in plain antero-posterior (AP) radiograph at 36 months post dual-level L3/L4 & L4/L5 Charitè disc replacement. Radiograph demonstrated circumferential formation (CF) of peri-prosthetic bone at L4/L5 (Lat radiograph not shown). Unilateral bridging (UB) and marginal osteophytes (O) were visible at L3/L4.

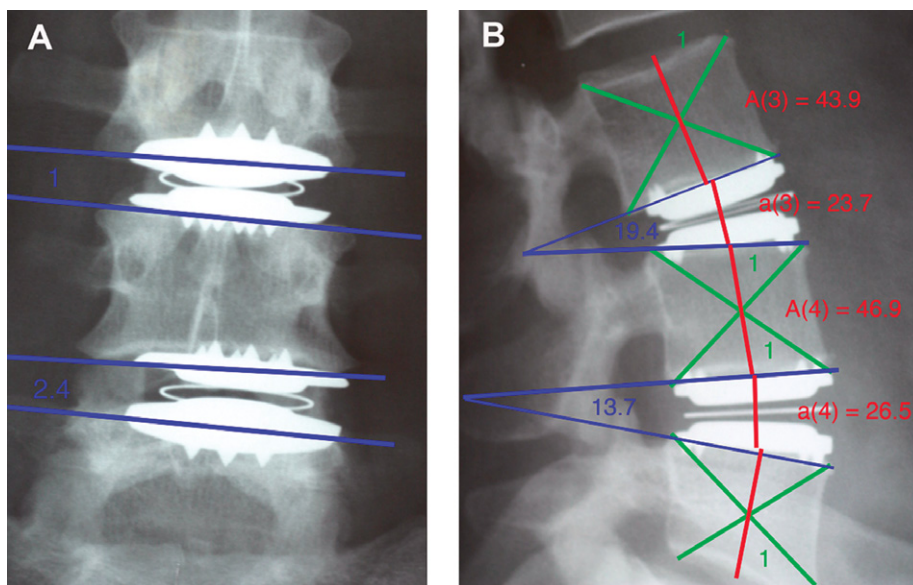


Fig. 3 Disc height restoration coefficient and angle defined stability method demonstrated in antero-posterior (A) and lateral (B) views following dual L3/L4 & L4/5 Charitè disc replacement. Disc height restoration coefficient was calculated by Equation 1.

Results

Twenty-five patients (13 females aged 33–59; 12 males aged 34–59) were identified for study.

Follow-up & Patient Characteristics

Following the initial attempt to contact all 25 patients ($n = 25$) with information packs and SF-36v2 health surveys, only two patients were responsive (8%). We were then able to contact 22 patients by phone and arrange clinical and radiologic assessment dates at our centre (12% lost to follow-up). Ten patients then attended for clinical assessment and self-completed their SF-36v2 health surveys (46% response rate). The remaining patients were contacted by phone, interviewed using the SF-36v2 health survey and sent a referral for radiographic assessment in their local area. Using this method, completed health surveys were obtained for all 22 contactable patients; however, radiographic assessment was available for 19 patients. Flexion/extension radiographs were obtained for three patients. In summary, from the initial 25 patients we attempted to contact, three (8%) were not contactable, 22 (92%) completed SF-36v2 health surveys and 19 (76%) were available for radiographic assessment. Mean postoperative follow-up was 34 months (standard deviation (SD) 14 months, maximum 64 months, and minimum 12 months). Mean BMI was 27.2 (± 4.7). Single-level disc replacements were performed in 15 patients and dual-level disc replacements were performed in four patients. Preoperative pain was typically low back pain and sciatic (18), with additional numbness (2), bladder/bowel dysfunction (1) or hip (1) pain. Pain was noted to have either begun for no reason (8), occurred at work (8), after surgery (2), after an illness (1), after a motor vehicle accident (1) or was of unspecified origin (2). Average duration of preoperative pain was 6.8 years (± 5.7).

Surgical Technical Accuracy

Radiographically confirmed single-level disc replacements were performed in 15 patients and dual-level disc replacements were performed in four patients. Ideal surgical placement was achieved overall in five (33%) single-level and three (37.5%) dual-level disc replacements. Sub-optimal surgical placement was seen in nine (60%) single-level and five (62.5%) dual-level disc replacements. An exact breakdown of surgical technical accuracy relative to the precise level of disc replaced for both single and dual-level disc replacements is presented in Table 2.

Disc Height Restoration Coefficient

Preoperative radiographs were provided by the 10 patients who attended for clinical assessment. Comparison of these radiographs to current radiographs enabled the disc height restoration (DHR) coefficient to be calculated. Mean disc height restoration achieved was 78.3%. A demonstrated increase in disc height of less than 50% was achieved in two patients, between 50% and 100% was achieved in three patients and greater than 100% was achieved in a further three patients. A decrease in disc height was illustrated in two patients. No statistically significant differences in DHR were found when

TABLE 2 Surgical technical accuracy vs level of disc replacement

Level of disc replacement	Surgical accuracy		
	Ideal	Sub-optimal	Poor
Single-level			
L3/L4	1	—	—
L4/L5	3	2	1
L5/S1	1	7	—
Dual-level			
L2/L3	—	1	—
L3/L4	1	1	—
L4/L5	1	2	—
L5/S1	1	1	—

comparing data grouped according to surgical technical accuracy, level of disc replacement, heterotopic ossification or BMI.

Angle Defined Stability

The stability of disc replacements (as defined by intervertebral angle $>10^\circ$) was assessed, with two (9%) joints radiographically determined to be unstable following disc replacement. No statistically significant difference in stability was seen when data were grouped for surgical accuracy, level of disc replacement or BMI. However, all instances of unstable intervertebral joints accompanied cases of apparently circumferential formation of peri-prosthetic bone.

Objective Pain Score Reduction

All patients demonstrated a reduction in objective pain score following treatment with Charitè disc replacement. Objective pain score reduction was statistically superior for single-level replacement in comparison to dual-level replacement ($P < 0.05$) (Fig. 4). Statistically significant differences in objective pain reduction were seen between data grouped according to the level of disc replacement, with L5/S1 showing greater pain reduction compared to L3/L4 ($P = 0.04$) (Fig. 5). Likewise, differences in objective pain score reduction between data grouped for heterotopic ossification, demonstrated that cases with osteophyte formation, unilateral bridging or bilateral bridging were statistically superior to joints with no formation of peri-prosthetic bone (Fig. 6).

SF-36v2 Outcomes

As with objective pain improvement, all SF-36v2 outcomes (except SF-36v2 *Mental Health*) were superior in single-level replacements compared to dual-level replacements (Fig. 7). Likewise, for ideal surgical accuracy all SF-36v2 outcomes were superior in comparison to the sub-optimal surgical placement group (Fig. 8). When compared to Australian population norms, SF-36v2 outcomes were superior for ideal placements

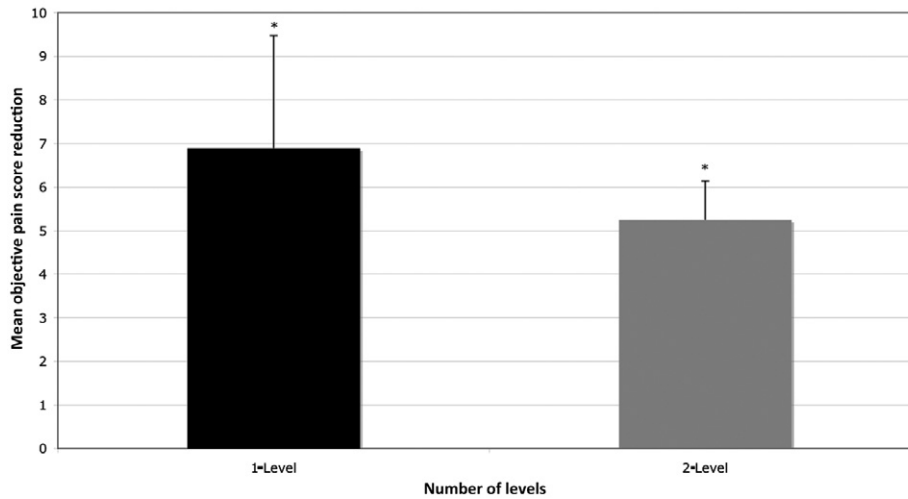


Fig. 4 Objective pain score reduction compared for data grouped according to the number of discs replaced. Pain score reductions were statistically superior for single-level replacement in comparison to dual-level replacement ($*P < 0.05$).

(Fig. 8). Statistically significant differences were demonstrated between osteophytes and bilateral bridging (Fig. 9). No statistically significant differences were seen for SF-36v2 outcomes when data were grouped according to the level of disc replacement or BMI weight classification.

Summary

All statistically significant differences in objective pain score reduction and SF-36v2 score between the grouping variables of surgical technical accuracy, number of levels replaced, level of

disc replacement, heterotopic ossification, patient BMI and DHR were summarized (Table 3).

Return-to-Work

Prior to surgery, all patients were either unemployed due to pain or undertaking part-time limited duties. As a practical measure, return to previous work was achieved in 89% of single-level disc replacement patients (11 full-time, five part-time), and 75% of dual-level disc replacement patients (one full-time, two part-time).

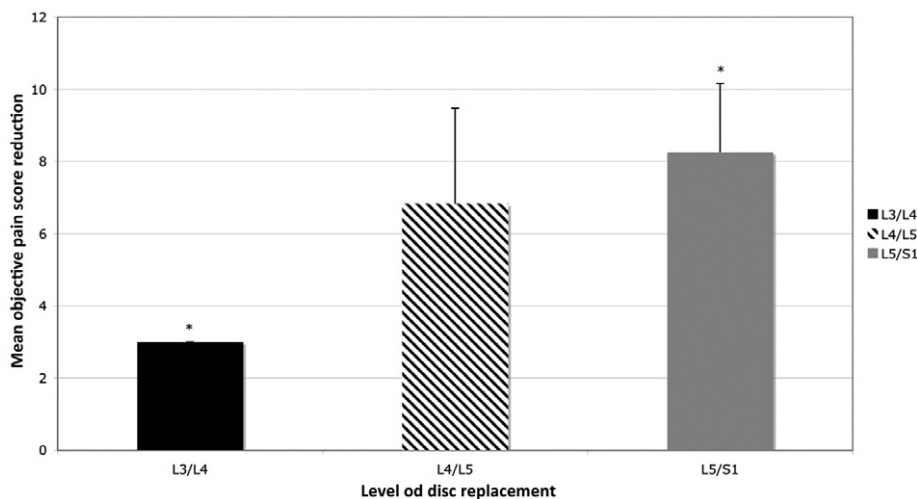


Fig. 5 Objective pain score reduction compared for data grouped according to the level of disc replacement. Statistically significant superior reductions demonstrated for L5/S1 compared to L3/L4 ($*P < 0.05$).

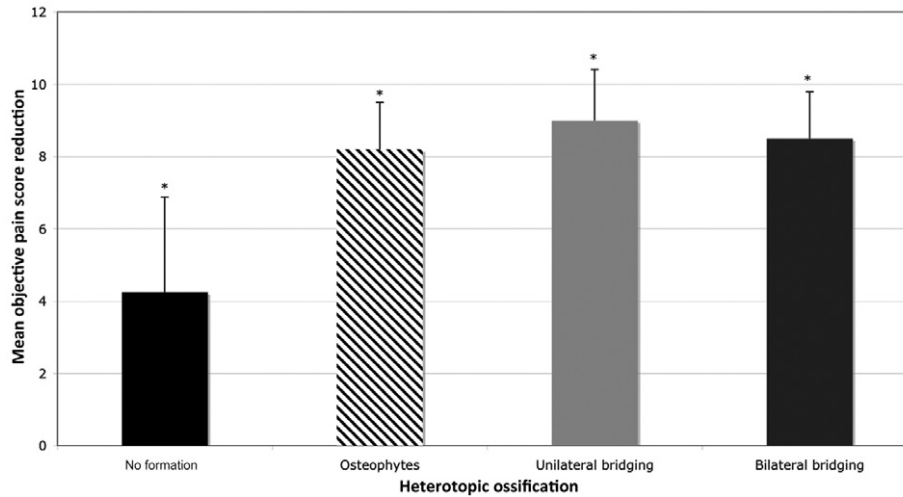


Fig. 6 Objective pain score reduction compared for data grouped according to the heterotopic ossification classification. Cases with osteophyte formation, unilateral bridging or bilateral bridging were statistically superior to joints with no formation of peri-prosthetic bone ($*P < 0.05$).

Discussion

All patients demonstrated a reduction in objective pain score following treatment with Charitè disc replacement. Objective pain score reduction was statistically superior for single-level replacement in comparison to dual-level replacement ($P < 0.05$). For the success of total disc replacement, adherence to the correct indications for the procedure and accurate surgical technique are considered paramount⁸. In our study, the accuracy of surgical placement was demonstrated to correlate to the clinical and functional outcomes, with ideal surgical placements statistically superior to sub-optimal place-

ments in SF-36v2 outcomes (Fig. 8). Reports remain limited concerning the direct correlation of radiological appearance, surgical technical accuracy and clinical outcomes, with one study previously addressing this topic; McAfee *et al.* (2005); the results of which are comparable to ours⁶. In Part II of the FDA regulated, IDE clinical trial, McAfee *et al.* (2005) evaluated both radiographic outcomes and the correlation of surgical technical accuracy to the clinical outcomes of Charitè total disc replacement using the Oswestry Disability Index (ODI) and Visual Analogue Scale (VAS) measures. For all patients in that study ($n = 276$), the surgical technical accuracy outcomes

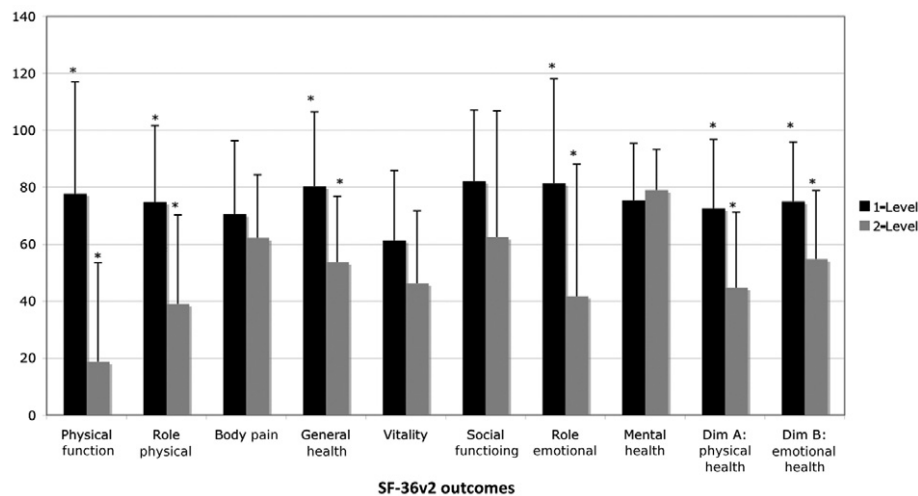


Fig. 7 Short Form 36-Health Survey (SF-36v2) outcomes compared for data grouped according to the number of discs replaced. All SF-36v2 outcomes (except SF-36v2 Mental Health) were superior in single-level replacements compared to dual-level replacements: statistically significant for Physical Function, Role Physical, General Health, Role Emotional, Physical Health, and Emotional Health ($*P < 0.05$).

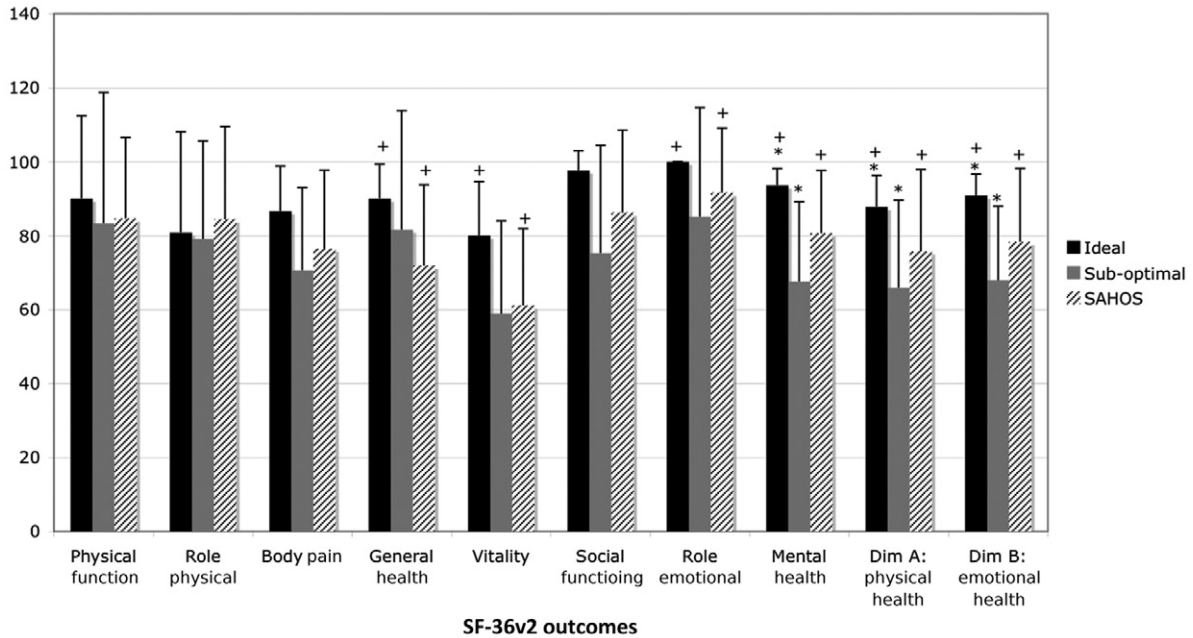


Fig. 8 Individual and overall Short Form 36-Health Survey (SF-36v2) outcomes compared to SF-36v2 Australian population norms (SAHOS) for data grouped according to surgical technical accuracy. All SF-36v2 outcomes were superior for ideal surgical accuracy in comparison to the sub-optimal surgical placement group; statistically significant for Mental Health, Physical Health, and Emotional Health (* $P < 0.05$). SF-36v2 outcomes were superior for ideal placements compared to SAHOS; statistically significant for General Health, Vitality, Role Emotional, Mental Health, Physical Health and Emotional Health (+ $P < 0.05$). Data sourced from: Hawthorne G, Osborne RH, Taylor A, and Sansoni J. The SF36 Version 2: critical analyses of population weights, scoring algorithms and population norms. *Quality of Life Research*, 2007, 4: 661–673.

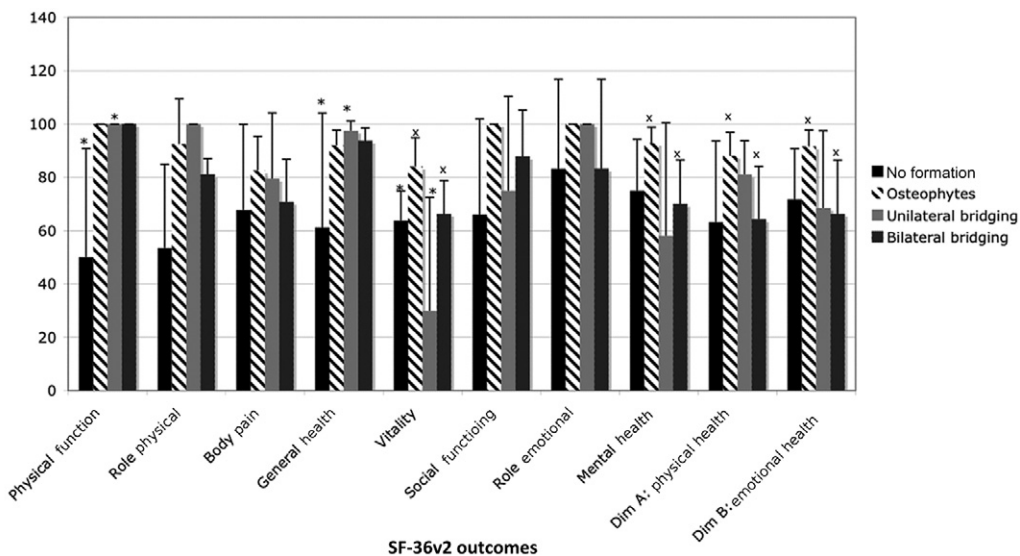


Fig. 9 Short Form 36-Health Survey (SF-36v2) outcomes compared for data grouped according to the heterotopic ossification classification. Statistically significant differences were shown between no formation and unilateral bridging for General Health, Physical Function and Vitality (* $P < 0.05$) and between osteophyte formation and bilateral bridging for Vitality, Mental Health, Physical Health, and Emotional Health (x $P < 0.05$).

TABLE 3 Summary of statistically significant differences (t-test; $P < 0.05$)

	# Levels	Level	Surgical accuracy	HO	BMI	DHR	Stability
Objective pain score	★	★	—	★	—	—	—
SF-36v2	★	—	★	★	—	—	—

★ = statistically significant differences, $P < 0.05$. BMI, body mass index; DHR, disc height restoration coefficient.

were: ideal (83%); sub-optimal (11%); and poor (6%), and correlation of surgical technical accuracy to both clinical outcome scales was shown to be statistically significant⁶. These results differ considerably from ours: ideal (33%), sub-optimal (60%), poor (7%); perhaps due to the stricter assessment criteria we used. However, as McAfee *et al.* (2005), our study also demonstrated improved outcomes associated with ideal surgical placements. These results confirm our initial hypothesis.

The clinical outcomes of lumbar disc replacement with the Charitè prosthesis have been examined by a number of previous studies, although never for an Australasian population³⁻⁷. The improvement in functional outcomes identified in our study are consistent with those of McAfee *et al.* (2003), who examined the clinical outcomes of Charitè disc replacement in contrast to the leading alternatives (lumbar fusion-interbody fusion cage and Bone Matrix Protein (BMP) or interbody autograft and pedicle screw instrumentation)⁷. Likewise, in Part I of the FDA regulated IDE clinical trial, Blumenthal *et al.* (2005) evaluated the clinical outcomes of Charitè using the SF-36 Health Survey, the VAS and the ODI as clinical outcome measures⁴. Patients in the Charitè group were shown to have lower levels of disability at all time intervals except the 24-month follow-up time-point⁴. Although including the SF-36 health survey in its methodology, no mention of the outcomes associated with this measure are presented in the results of Blumenthal *et al.* (2005). To our knowledge, our study is therefore the first to present the results of SF-36v2 outcomes for a group of patients receiving Charitè lumbar disc replacement in Australia and worldwide.

In our study, the factors affecting the degree of surgical accuracy achieved were also considered, with no statistically significant differences in surgical technical accuracy seen between data grouped according to patient BMI or the level of disc placement. However a trend suggested that sub-optimal placement is more likely at the L5/S1 level, a finding consistent with our practical surgical experience. Blumenthal *et al.* (2005) determined that BMI (and gender and age) had no impact on clinical outcomes⁴. This was confirmed in our study; however, there was a substantial difference in mean BMI between patients receiving single- and dual level disc replacements, although this was not statistically significant.

Our experience with return to work following surgery (89% single-level, 75% dual-level) is consistent with Lemaire *et al.* (2005) who reported a return to full-time or part-time work was achieved in 89% of patients³. In that series, Lemaire

et al. (2005) also reported a mean increase in interspace height of 51.3%, with only a single case of decreased interspace height (thought to suggest polyethylene wear)³. This increase in interspace height is significantly less than the mean increase in disc height restoration (DHR) of 78.3% in our series, perhaps due to relatively later surgical intervention than in the European population. Despite these differences, our radiographic assessment technique offers a number of distinct advantages: due to the purely geometric nature, the influence of inter-observer variability is minimized and the need for radiographic scaling is obviated.

McAfee *et al.* (2003) has previously addressed the classification of heterotopic ossification (HO) in lumbar disc replacement¹³. In an effort to further analyze the relationship between heterotopic ossification and clinical outcomes, our study did not use the classification system as used by McAfee, rather, we assessed and compared the variables independently (no formation, osteophyte formation, unilateral bridging, bilateral bridging or circumferential formation). Using data from the IDE trial, Tortolani *et al.* (2007) have recently used McAfee's classification system to examine the issue of heterotopic ossification around the Charitè disc replacement in relation to both range of motion (ROM) and clinical outcomes¹⁴. In that study, the prevalence of ossification was found to be 4.3% ($n = 276$) with four cases of Class-I and eight cases of Class-II. Tortolani *et al.* (2007) determined that the presence of heterotopic ossification did not affect ROM or clinical outcomes in their patient group at 24-month follow-up¹⁴. The incidence of heterotopic ossification was also reported by Lemaire *et al.* (2005) at 3%; David (2002) at 7.7%; and Marnay (2002) at 1.7%^{3,15,16}. Our results demonstrated no formation of peri-prosthetic bone in 17% of disc replacements, marginal osteophytes in 39%, unilateral bridging in 17%, bilateral bridging in 17% and apparently circumferential formation in 9% of disc replacements. Given our results, Tortolani's conclusion that the occurrence of "heterotopic ossification is infrequent following lumbar total disc replacement with the Charitè artificial disc" does not apply to our patient group. In contrast to previous studies, our results demonstrate a degree of association between heterotopic ossification and clinical outcomes^{3,15,16}.

Limitations

There are a number of limitations to the study described herein: the relatively small sample size, the lack of preoperative

baseline data and the limited availability of dynamic flexion/extension radiographs. The small sample size was restrictive in that it prevented the use of more powerful analysis methods such as the Kruskal–Wallis test; however, significant differences between group means were shown using a standard Student's *t*-test. It must also be acknowledged that the number of disc replacements undertaken annually in Australia remains limited; as such this study makes a significant contribution to our knowledge of how these implants perform from a single centre within Australia.

As stated in the methods, a lack of baseline data dictated that preoperative pain scores be obtained retrospectively. While not an ideal study design, Dawson *et al.* (2002) suggest that retrospective study of the outcome of spinal surgery patients can deliver valuable information providing the limitations are considered¹⁷. Factors such as patient age, gender, reporting medium and the current status of pain, may influence the accuracy of retrospective scores^{18,19}. It has also been shown that a patient's memory of their health status can be altered by experiences that have occurred in the postoperative interval and that data may further be influenced by patient beliefs regarding the effectiveness of treatment^{20,21}. Despite these limiting factors, retrospective study designs are common throughout the spinal surgical research literature; a review of MEDLINE reported a greater rate of increase in retrospective

studies (300%) compared to either randomized controlled trials (80%) or prospective studies (160%)¹⁷. Interestingly, Dawson *et al.* (2002) point out that a comparable reliance upon patient memory and recall is ubiquitous throughout the course of history taking in the medical and surgical setting¹⁷. It has been evidenced that, as used in our study, the recall of general pain intensity is good; however, the specific qualities of that pain may not be as well recalled^{22,23}. Also, a recent report in the palliative care literature suggests that retrospective pain scoring may be a sufficiently reliable method for research purposes²⁴. Ultimately though, a reliance on retrospective pain data in cross-sectional studies may lead to an overestimation of the effectiveness of surgery²¹. Ergo in our study, a complete dependence on retrospective data was avoided by comparing data to a validated baseline of previously established Australian SF-36v2 population norms.

Lumbar disc replacement with the Charité prosthesis resulted in retrospective objective pain score reduction for all patients examined in this study. When grouped according to radiographically assessed surgical technical accuracy, ideal placements were associated with improved SF-36v2 functional outcomes in comparison to sub-optimal placements. This result confirms the hypothesis that ideal surgical placements are associated with improved outcomes in total lumbar disc replacement.

References

1. Bao QB, Yuan HA. Prosthetic disc replacement: the future? *Clin Orthop Relat Res*, 2002, 394: 139–145.
2. German JW, Foley KT. Disc arthroplasty in the management of the painful lumbar motion segment. *Spine*, 2005, 30 (16 Suppl.): S60–S67.
3. Lemaire JP, Carrier H, Sariali H, *et al.* Clinical and radiological outcomes with the Charité artificial disc: a 10-year minimum follow-up. *J Spinal Disord Tech*, 2005, 18: 353–359.
4. Blumenthal S, McAfee PC, Guyer RD, *et al.* A prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion: part I: evaluation of clinical outcomes. *Spine*, 2005, 30: 1565–1575.
5. Blumenthal SL, Ohnmeiss DD, Guyer RD, *et al.* Prospective study evaluating total disc replacement: preliminary results. *J Spinal Disord Tech*, 2003, 16: 450–454.
6. McAfee PC, Cunningham B, Holsapple G, *et al.* A prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion: part II: evaluation of radiographic outcomes and correlation of surgical technique accuracy with clinical outcomes. *Spine*, 2005, 30: 1576–1583.
7. McAfee PC, Fedder IL, Saiedy S, *et al.* SB Charité disc replacement: report of 60 prospective randomized cases in a US center. *J Spinal Disord Tech*, 2003, 16: 424–433.
8. Geisler FH. Surgical technique of lumbar artificial disc replacement with the Charité artificial disc. *Neurosurgery*, 2005, 56 (1 Suppl.): 46–57.
9. Kuslich SD, Ulstrom CL, Griffith SL, *et al.* The Bagby and Kuslich method of lumbar interbody fusion. History, techniques, and 2-year follow-up results of a United States prospective, multicenter trial. *Spine*, 1998, 23: 1267–1279.
10. Mochida J, Nishimura K, Nomura T, *et al.* The importance of preserving disc structure in surgical approaches to lumbar disc herniation. *Spine*, 1996, 21: 1556–1563.
11. Ware JE, Kosinski M, Dewey JE. How to Score Version 2 of the SF-36 Health Survey, 3rd edn. Lincoln, RI: QualityMetric Incorporated, 2001: 238.
12. Hawthorne G, Osborne RH, Taylor A, *et al.* The SF36 Version 2: critical analyses of population weights, scoring algorithms and population norms. *Qual Life Res*, 2007, 16: 661–673.
13. McAfee PC, Cunningham BW, Devine J, *et al.* Classification of heterotopic ossification (HO) in artificial disk replacement. *J Spinal Disord Tech*, 2003, 16: 384–389.
14. Tortolani PJ, Cunningham BW, Eng M, *et al.* Prevalence of heterotopic ossification following total disc replacement. A prospective, randomized study of two hundred and seventy-six patients. *J Bone Joint Surg Am*, 2007, 89: 82–88.
15. David T. Lumbar disc prosthesis: five years follow-up study on 147 patients with 163 SB CHARITE prosthesis. The Fourth Annual Meeting of Spine Society of Europe, 2002; Nantes, France.
16. Marnay T. Lumbar disc replacement 1–11 years results with ProDisc. The Fourth Annual Meeting of Spine Society of Europe, 2002; Nantes, France.
17. Dawson EG, Kanim LE, Sra P, *et al.* Low back pain recollection versus concurrent accounts: outcomes analysis. *Spine*, 2002, 27: 984–994.
18. Boyer GS, Templin DW, Goring WP, *et al.* Discrepancies between patient recall and the medical record. Potential impact on diagnosis and clinical assessment of chronic disease. *Arch Intern Med*, 1995, 155: 1868–1872.
19. Pincus T, Pearce S, McClelland A, *et al.* Self-referential selective memory in pain patients. *Br J Clin Psychol*, 1993, 32 (Pt 3): 365–374.
20. Lingard EA, Wright EA, Sledge CB, *et al.* Pitfalls of using patient recall to derive preoperative status in outcome studies of total knee arthroplasty. *J Bone Joint Surg Am*, 2001, 83-A: 1149–1156.
21. Pellisé F, Vidal X, Hernández A, *et al.* Reliability of retrospective clinical data to evaluate the effectiveness of lumbar fusion in chronic low back pain. *Spine*, 2005, 30: 365–368.
22. Beese A, Morley S. Memory for acute pain experience is specifically inaccurate but generally reliable. *Pain*, 1993, 53: 183–189.
23. Niven CA, Brodie EE. Memory for labor pain: context and quality. *Pain*, 1996, 64: 387–392.
24. Brauer C, Thomsen JF, Loft IP, *et al.* Can we rely on retrospective pain assessments? *Am J Epidemiol*, 2003, 157: 552–557.