

Comparison of distribution, agreement and correlation between the original and modified Merle d'Aubigné-Postel Score and the Harris Hip Score after acetabular fracture treatment

Moderate agreement, high ceiling effect and excellent correlation in 450 patients

Stein Øvre¹, Leiv Sandvik², Jan Erik Madsen¹ and Olav Røise¹

¹Orthopaedic Centre, Ullevål University Hospital, NO-0407 Oslo, ²Research Forum, University of Oslo, NO-0407 Oslo, Norway
Correspondence SØ: SteinA.Ovre@ullevaal.no
Submitted 04-09-09. Accepted 05-02-15

Background In acetabular fracture treatment, 3 disease-specific outcome scores are mainly used: the original and modified Merle d'Aubigné-Postel Score, and the Harris Hip Score.

Methods The original and modified Merle d'Aubigné-Postel Score and the Harris Hip Score were recorded in 1,153 follow-ups of 450 patients. 492 follow-ups were excluded because factors other than the acetabular fracture were found to affect the outcome scores. This gave 661 patient records for the study.

Results The Spearman correlations were between 0.81 and 0.89. The quartile analyses showed Kappa agreement between 0.45 and 0.55. About 40% of the observations were classified into another quartile when switching from one outcome score to another. The 25th and 50th percentiles comprised 85% and 95% of the total numeric scores, respectively, while the 75th percentiles showed ceiling value (100% of the maximum) in all 3 scores.

Interpretation Despite the excellent overall correlation between the outcome scores, the Kappa agreements were only moderate. The scores were all skewed in distribution with considerable ceiling effects that could limit their clinical use. The scores did not capture any differences in 25% of the observations at the upper end of the scales.

In acetabular fracture treatment, 3 disease-specific outcome scores are mainly used. The Merle d'Aubigné and Postel Score (1954) (Table 1) was originally used to evaluate functional results in hip arthroplasties. Letournel and Judet (1993) used this score for assessment of the results of acetabular fracture treatment. The Merle d'Aubigné-Postel Score was modified by Matta et al. (1986) (Table 2). The clinical grading was directly related to the total numeric outcome score. Harris introduced his much used Hip Score in 1969 as a method of evaluating treatment with mold arthroplasties of the hip, secondary to arthrosis after dislocation and acetabular fractures (Harris 1969, Helfet et al. 1992, Mears et al. 2003) (Table 3).

An instrument for measurement of clinical outcome should give information about patient function and pain, and it is crucial that it should be both valid and reliable (Pynsent 2001). Comparisons between the Harris Hip Score and the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36) (Shields et al. 1995) have indicated that the Harris Hip Score is a suitable instrument for evaluation of changes in hip function (Hoeksma et al. 2003). To a certain extent, it captures quality-of-life domains influenced by total hip arthroplasties (Lieberman et al. 1997). In addition, a correlation has been described between the modified Merle d'Aubigné-Postel Score and the Musculoskeletal

Table 1. Original Merle d'Aubigné-Postel Score. The individual scores of Pain, Walking ability and Mobility are added together to give an overall numeric score. Clinical grades (Very good, Good, Medium, Fair, Poor) are given by the scores of Pain and Walking ability and adjusted down 1–2 grades, depending on the mobility score (D'Aubigne and Postel 1954). The total numeric score was used in our study

Score	Pain	Walking ability	Mobility
0	Pain is intense and permanent	None	Ankylosis with bad position of the hip
1	Pain is severe, even at night	Only with crutches	No movement; pain or slight deformity
2	Pain is severe when walking; prevents any activity	Only with canes	Flexion under 40 degrees
3	Pain is tolerable with limited activity	With 1 cane, less than 1 h; very difficult	Flexion between 40 and 60 degrees
4	Pain is mild when walking; it disappears with rest	A long time with a cane; short time without cane and with limp	Flexion between 60 and 80 degrees; patient can reach his/her foot
5	Pain is mild and inconstant; normal activity	Without cane, but with slight limp	Flexion between 80 and 90 degrees; abduction of at least 15 degrees
6	No pain	Normal	Flexion more than 90 degrees; abduction to 30 degrees

Table 2. Modified Merle d'Aubigné-Postel Score. The overall numeric score is given by adding the domain scores. Clinical grades: Excellent 18, Good 15–17, Fair 12–14, Poor 3–11 (Matta et al. 1986). The total numeric score was used in the present study

Score	Pain	Ambulation	ROM (%)
6	No pain	Normal	95–100
5	Slight or intermittent	No cane, but slight limp	80–94
4	Pain after ambulation, but disappears	Long distances with cane or crutch	60–79
3	Moderately severe, permits ambulation	Limited, even with support	40–59
2	Severe with ambulation	Very limited	
1	Severe, prevents ambulation	Bedridden	0–39

Function Assessment health-status questionnaire (Engelberg et al. 1999, Swiontkowski et al. 1999, Saterbak et al. 2000, Moed et al. 2003) However, the correlations and the differences between the original and the modified Merle d'Aubigné-Postel Scores, and also the Harris Hip Score, are unknown. We compared these 3 outcome scores.

Patients and methods

All patients with an acetabular fracture who were admitted to Ullevål University Hospital between 1993 and 2003 were recorded prospectively. The fractures were classified according to Letournel and Judet (1993). Follow-ups at 6 and 12 months and at 2, 5, 7–8 and 10 years were recorded. At each follow-up, radiographs were taken in addition to the following clinical outcome scores: 1) the original Merle d'Aubigné-Postel Score (D'Aubigne and

Postel 1954), 2) the modified Merle d'Aubigné-Postel Score (Matta et al. 1986), and 3) the Harris Hip Score (Harris 1969). Follow-ups influenced by factors other than the acetabular fracture were excluded in order to minimize bias.

450 patients were recorded, with 1,153 follow-ups. The median age was 44 (12–92) years. 53 patients died because of the initial trauma or during follow-up. Exclusion of biased records gave 661 observations to include in the study, where only the acetabular fracture would affect the outcome scores.

Original Merle d'Aubigné-Postel Score

The Merle d'Aubigné-Postel Score gives individual scores for the domains Pain, Mobility and Walking Ability (Table 1). The scores for pain and ability to walk are added and then classified into the grades very good, good, medium, fair and poor. These gradings are then adjusted down by

Table 3. The Harris Hip Score consists of the domains Pain, Function, Deformity and ROM, and gives a maximum of 100 points in the total score (Harris 1969)

Domain	Score
I Pain	
A None or ignores it	44
B Slight, occasional, no compromise in activities	40
C Mild pain, no effect on average activities, rarely moderate pain with unusual activity, may take aspirin	30
D Moderate pain, tolerable but makes concessions to pain. Some limitation of ordinary activity or work. May require occasional pain medicine stronger than aspirin	20
E Marked pain, serious limitations of activities	10
F Totally disabled, crippled, pain in bed, bedridden	0
II Function	
A Gait	
1 Limp:	
a None	11
b Slight	8
c Moderate	5
d Severe	0
2 Support:	
a None	11
b Cane for long walks	7
c Cane most of the time	5
d One crutch	3
e Two canes	2
f Two crutches	0
g Not able to walk	0
B Activities	
1 Stairs:	
a Normally without using any railing	4
b Normally using a railing	2
c In any manner	1
d Unable to use stairs	0
2 Shoes and socks:	
a With ease	4
b With difficulty	2
c Unable	0
3 Sitting:	
a Sitting comfortably in ordinary chair 1 h	5
b On a high chair for one-half hour	3
c Unable to sit comfortably in any chair	0
4 Distance:	
a Unlimited	11
b Six blocks	8
c Two or three blocks	5
d Indoors only	2
e Bed and chair	0
5 Enter public transportation	1
III Absence of deformity points (4) are given if the patient demonstrates:	
A Less than 30° fixed flexion contracture	
B Less than 10° fixed adduction	
C Less than 10° fixed internal rotation in extension	
D Limb-length discrepancy less than 3.2 cm	
IV ROM (Index values are determined by multiplying the degrees of motion by the indices. The ROM score is given by adding the index values multiplying with the factor 0.05. Maximum score gives 5 points)	
A Flexion: 0–45° × 1.0, 45–90° × 0.6, 90–110° × 0.3	
B Abduction: 0–15° × 0.8, 15–20° × 0.3, > 20° × 0	
C External rotation in extension: 0–15° × 0.4, > 15° × 0	
D Internal rotation in extension: any × 0	
E Adduction: 0–15° × 0.2	

1–2 grades for the mobility score to give the final clinical grade (D'Aubigne and Postel 1954). In this study we used the total numeric score as the basis for statistics, and thus the three domains in the Merle d'Aubigné-Postel Score have the same impact (one-third each; Table 1).

Modified Merle d'Aubigné-Postel Score

The modified Merle d'Aubigné-Postel Score (Matta et al. 1986) is based on the same components as the original and individual scores, but with a slight difference in language and grading (Table 2). The pain and ambulation domains are split into 6 grades (1–6, and not 0–6 as originally described). The main difference is in the domain mobility/ROM where the score is related to relative ROM and split into 5 grades with points given as follows: 1, 3, 4, 5 and 6. The clinical grading relates directly to the total numerical score. Again, the three domains have an impact of one-third each in the numerical outcome.

Harris Hip Score

The Harris Hip Score includes the domains of Function, Pain, Motion and Deformity (Table 3). It is a disease-specific instrument originally intended as an outcome score after mold arthroplasties. The maximum score is 100 points and the Pain domain contributes 44 points, Function 47, ROM 5 and Absence of Deformity 4 points. Function is subdivided into gait and activities of daily living. Calculation of the ROM score includes splitting the motion into categories based on utility and then multiplying the degrees of motion with a given index factor. The index scores are then added and multiplied by a factor of 0.05 to obtain the final ROM score.

Statistics

Statistical analysis was performed with NCSS Statistical Software for Windows (version 2004; NCSS and PASS, Kaysville, Utah). The statistical analyses were based on the total numeric score and not on clinical grades. Descriptive statistics and Kappa agreement analysis was evaluated according to the following guidelines: poor agreement (Kappa value < 0.20), fair (0.21–0.40), moderate (0.41–0.60), good (0.61–0.80) and very good (0.81–1.00) (Altman 1991). We performed the

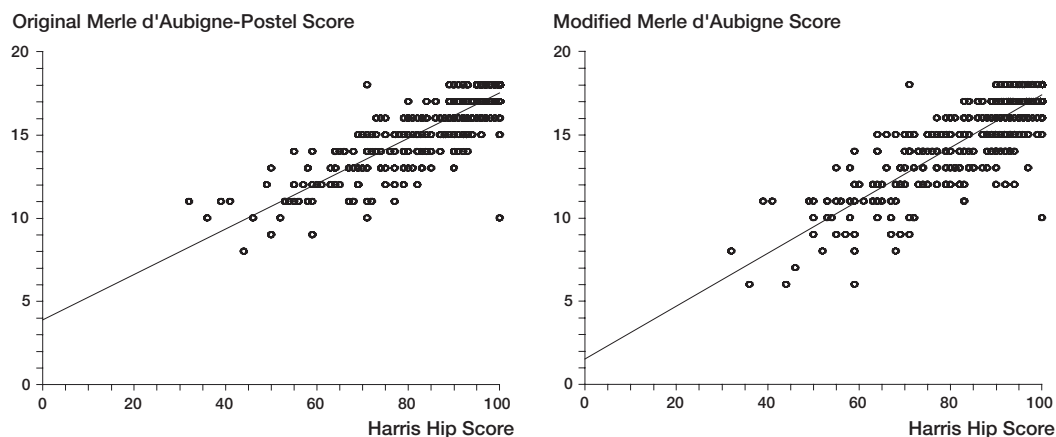


Figure 1. The scatter plots between total numeric score of the Harris Hip Score and the original- and modified Merle d'Aubigné-Postel Scores show how well the scores conform to a straight line. Excellent correlations are apparent, but a defined value in one score may represent a considerable variation in the other score.

agreement analysis on the total outcome of the three scores. Regression analysis with Spearman's correlation coefficient was done on the total and domain scores. There is no correlation when the coefficient is 0, whereas a value of 1 shows a perfect positive correlation. Spearman's correlation coefficient was interpreted according the following guidelines: poor ($r < 0.3$), moderate ($0.3 < r < 0.6$), good ($0.6 < r < 0.8$) and excellent correlation ($r > 0.8$) (Bellamy et al. 1991).

Results

Overall correlation and agreement

The mean original Merle d'Aubigné-Postel Score was 16 (8–18), SD 2.0, and the mean modified Merle d'Aubigné-Postel Score was 15 (8–18), SD 2.4. The mean Harris Hip Score was 90 (32–100), SD 13.

The Spearman correlation between Harris Hip Score and the original Merle d'Aubigné-Postel Score was $r = 0.82$. Between Harris Hip Score and the modified Merle d'Aubigné-Postel Score the correlation was $r = 0.81$, and between the original and modified Merle d'Aubigné-Postel Score the correlation was $r = 0.89$ (Table 4). Scatter plots with the regression lines of Harris Hip Score and the Merle d'Aubigné-Postel Scores are shown in Figure 1.

The Kappa agreement between Harris Hip Score and the original Merle d'Aubigné-Postel Score

was 0.49, and between the Harris Hip Score and the modified score the value was 0.45. The agreement between the original and modified Merle d'Aubigné-Postel Scores was 0.55. Shifts of observations (outcomes) from one quartile in one score to another quartile in another score were significant. Comparing the Harris Hip Score and the original and modified Merle d'Aubigné-Postel Scores, the total shift was about 40%. The Harris Hip Score could differ by more than 20 points for a given numeric score in the Merle d'Aubigné-Postel Scores. The variations between the Merle d'Aubigné-Postel Scores could be more than 5 points.

Analysis of the pain scores

The original and modified Merle d'Aubigné-Postel Pain Scores were identical, with mean 5.2 (2–6), SD 0.96. The mean Pain score of the Harris Hip Score was 40 (10–44), SD 5.7. Consequently, the correlation was equal between the Pain score of the Harris Hip Score and the original and modified Merle d'Aubigné-Postel Pain score, with $r = 0.81$ (Table 4).

Analysis of the mobility/ROM scores

The mean of the original Merle d'Aubigné-Postel Mobility score was 5.6 (4–6), SD 0.56, whereas the mean of the modified Merle d'Aubigné-Postel ROM score was 5.2 (1–6), SD 1.0. The mean Harris Hip score was 4.9 (2–5), SD 0.39. The coef-

Table 4. Spearman correlation coefficients between the domains in the three different outcome scores

Harris Hip Score	0.82	Original Merle d'Aubigné-Postel Score	
Pain		0.81	Pain
Function		0.78	Walking ability
ROM		0.29	Mobility
Original Merle d'Aubigné-Postel Score	0.89	Modified Merle d'Aubigné-Postel Score	
Pain		1.0	Pain
Walking ability		1.0	Ambulation
Mobility		0.31	ROM relative
Modified Merle d'Aubigné-Postel Score	0.81	Harris Hip Score	
Pain		0.81	Pain
Ambulation		0.78	Function
ROM relative		0.33	ROM

coefficient of correlation between the Harris Hip Score ROM domain and the original Merle d'Aubigné-Postel Mobility score was 0.29, and the r value was 0.33 for the corresponding correlation involving the modified Merle d'Aubigné-Postel ROM score. The original and modified Merle d'Aubigné-Postel Pain scores had a correlation coefficient of 0.31 (Table 4).

Analysis of function scores

Again, the function scores of the original and modified Merle d'Aubigné-Postel Scores (Walking Ability and Ambulation) were identical, with mean 5.3 (1–6), SD 1.1. The Function score of the Harris Hip Score was 50 (11–56), SD 8.6. The coefficient of correlation between the Function score of the Harris Hip Score and the Merle d'Aubigné-Postel Function score was 0.78 (Table 4).

Division into operative and nonoperative treatment

The 661 records were divided into nonoperatively treated (278) and operatively treated (383) groups. The Spearman coefficient of correlation between original and modified Merle d'Aubigné-Postel Scores in the operated group was 0.91, and corresponding figure for the conservatively treated group was similar ($r = 0.85$). When comparing the correlation between Harris Hip Score and the original and modified Merle d'Aubigné-Postel Scores in the operated group, the coefficients were 0.84 and 0.83, respectively. For the nonoperatively treated group, the coefficients were almost the same (0.79 and 0.75, respectively).

Distribution

All 3 outcome scores were highly asymmetric (Figure 2), with a considerable ceiling effect, i.e. dominance of top scores. In the Harris Hip Score, the 25th and 50th percentiles showed the values of 86 and 96 points, respectively, while the 75th gave the ceiling value of 100 points. The original and modified Merle d'Aubigné-Postel Scores were almost identical, and showed the same values for the 25th and 50th percentiles (15 and 17 points, which represent 83% and 94% of the top score, respectively). The 75th percentile gave the ceiling value of 18 points.

Discussion

We found a close correlation between the original and modified Merle d'Aubigné-Postel Scores. The difference between these scores relates to a modification of the Mobility domain. In the original Merle d'Aubigné-Postel Score, the flexion and abduction ability are combined, while in the modified score, relative ROM constitutes this domain. The modified Merle d'Aubigné-Postel Score splits the domains of Pain and Ambulation/Walking Ability into 6 grades (points 1–6), as opposed to 7 (points 0–6) as originally described. This small difference in design did not affect the total numeric score in this study at all. The lowest grades (0 and 1) of the two changed domains in the original and modified score were not applicable to any patient in our study, and the other grades (2–6) in the two scores relate to almost identical situations. Out-

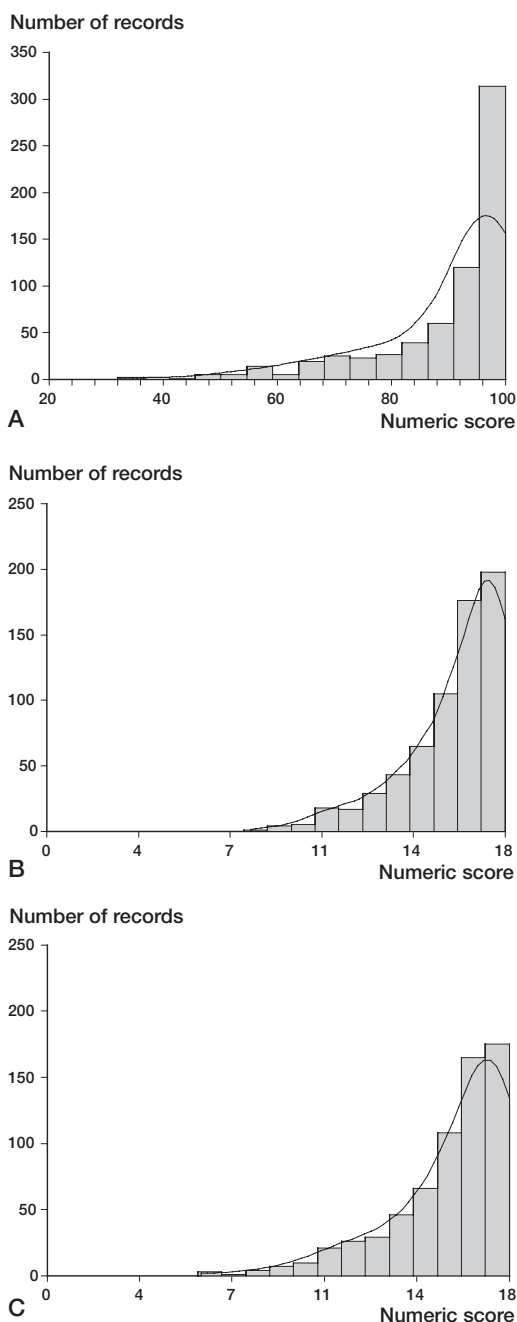


Figure 2. The distributions of the 661 records in the three outcome scores (A – Harris Hip Score, B – Original Merle d'Aubigné-Postel Score, and C – Modified Merle d'Aubigné-Postel Score) were skewed with a left tail and demonstrated a considerable ceiling effect for all three.

come studies of hip disabilities using the original or the modified Merle d'Aubigné-Postel Score will thus produce almost identical results.

The correlations between Harris Hip Score and the original and modified Merle d'Aubigné-Postel Scores were excellent according to the guidelines chosen (Bellamy et al. 1991). The original and modified Merle d'Aubigné-Postel Scores demonstrated almost identical correlations to the Harris Hip Score in this study. Even so, there was a slightly closer correlation between the Harris Hip Score ROM component and the relative ROM of the modified Merle d'Aubigné-Postel Score (compared to the Harris Hip Score and the original Merle d'Aubigné-Postel Score). The pain components compared between the Harris Hip Score and the Merle d'Aubigné-Postel Scores correlated less than the overall scores, but the correlations were still good. The Function domains of the Merle d'Aubigné-Postel Scores were identical in numeric scores and showed excellent correlation with the Function domain of the Harris Hip Score.

Dividing the selection into operative and non-operative treatments only affected the correlations between the outcome scores to a minor degree. The original and modified Merle d'Aubigné-Postel Scores and the Harris Hip Score assessed the patients almost identically, irrespective of kind of treatment. There was a small tendency of better correlation when treating patients operatively rather than conservatively, but overall, the scoring systems were almost equally sensitive to possible changes in outcome due to treatment.

Despite the excellent overall correlation between the outcome scores, the Kappa agreements were only moderate. About 40% of the observations were classified into another quartile when switching from one outcome score to another. This may affect the conclusions when comparing outcome with different scoring systems in smaller patient series.

The outcome scores were all skewed in distribution, with considerable ceiling effects. The ceiling in the original and modified Merle d'Aubigné-Postel Scores has also been demonstrated and discussed by others (Rice et al. 2002, Moed et al. 2003). We found the same phenomenon with the Harris Hip Score in the same study population. The most likely explanation would be that the instruments are skewed in design, or that the outcome of the patient population was rather good. Even though the domains in the outcome scores have

different effects on the total numeric scores, the ceiling effects are very similar when comparing the percentiles.

There was more than one observation of several patients in the study which could be misleading for the conclusions. However, with the large number of patients included, this was not critical for the statistics and, additionally, the purpose of the study was to evaluate the outcome scores and not the outcome of the total patient population. The large number of observations in the present study compensated for the moderate Kappa agreements and considerable shifts of observations between the quartiles. Despite the moderate Kappa values, the Spearman correlation coefficients were valid.

The ceiling phenomenon limits the clinical value of the outcome scores. A disease-specific outcome score should discriminate better between outcomes at the upper end of the scale, in order to make it easier to analyze factors affecting the treatment result. There should be further research on discrimination in disease-specific outcome scores of acetabular fracture treatment.

No competing interests declared.

Altman D G. Practical statistics for medical research. Chapman & Hall 1991.

Bellamy N, Wells G, Campbell J. Relationship between severity and clinical importance of symptoms in osteoarthritis. *Clin Rheumatol* 1991; 10 (2): 138-43.

D'Aubigne R M, Postel M. Functional results of hip arthroplasty with acrylic prosthesis. *J Bone Joint Surg (Am)* 1954; 36 (3): 451-75.

Engelberg R, Martin D P, Agel J, Swiontkowski M F. Musculoskeletal function assessment: reference values for patient and non-patient samples. *J Orthop Res* 1999; 17 (1): 101-9.

Harris W H. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. *J Bone Joint Surg (Am)* 1969; 51 (4): 737-55.

Helfet D L, Borrelli J, Jr., DiPasquale T, Sanders R. Stabilization of acetabular fractures in elderly patients. *J Bone Joint Surg (Am)* 1992; 74 (5): 753-65.

Hoeksma H L, Van Den Ende C H, Runday H K, Heering A, Breedveld F C. Comparison of the responsiveness of the Harris Hip Score with generic measures for hip function in osteoarthritis of the hip. *Ann Rheum Dis* 2003; 62 (10): 935-8.

Letournel E, Judet R. Fractures of the acetabulum. Springer-Verlag, New York, London, Berlin, Heidelberg 1993.

Lieberman J R, Dorey F, Shekelle P, Schumacher L, Kilgus D J, Thomas B J, Finerman G A. Outcome after total hip arthroplasty. Comparison of a traditional disease-specific and a quality-of-life measurement of outcome. *J Arthroplasty* 1997; 12 (6): 639-45.

Matta J M, Mehne D K, Roffi R. Fractures of the acetabulum. Early results of a prospective study. *Clin Orthop* 1986; (205): 241-50.

Mears D C, Velyvis J H, Chang C P. Displaced acetabular fractures managed operatively: indicators of outcome. *Clin Orthop* 2003; (407): 173-86.

Moed B R, Yu P H, Gruson K I. Functional outcomes of acetabular fractures. *J Bone Joint Surg (Am)* 2003; 85 (10): 1879-83.

Pynsent P B. Choosing an outcome measure. *J Bone Joint Surg (Br)* 2001; 83 (6): 792-4.

Rice J, Kalisz M, Dolan M, Cox M, Khan H, McElwain J P. Comparison between clinical and radiologic outcome measures after reconstruction of acetabular fractures. *J Orthop Trauma* 2002; 16 (2): 82-6.

Saterbak A M, Marsh J L, Nepola J V, Brandser E A, Turbett T. Clinical failure after posterior wall acetabular fractures: the influence of initial fracture patterns. *J Orthop Trauma* 2000; 14 (4): 230-7.

Shields R K, Enloe L J, Evans R E, Smith K B, Steckel S D. Reliability, validity, and responsiveness of functional tests in patients with total joint replacement. *Phys Ther* 1995; 75 (3): 169-76; discussion 176-9.

Swiontkowski M F, Engelberg R, Martin D P, Agel J. Short musculoskeletal function assessment questionnaire: validity, reliability, and responsiveness. *J Bone Joint Surg (Am)* 1999; 81 (9): 1245-60.