Original article

Comparison and agreement of outcome scores through nine months after acetabular fracture fixation

José Luis Ramirez-Garcialuna^{a,b,*}, Fernando Dominguez-Paulin^c,
Jesús Ramirez-Martinez^c, Luis Fernando Sanmiguel-Delgadillo^c

^a Experimental Surgery Program, Surgery Division of the Research Institute of the McGill University Health Centre, 1650 Cedar Ave, H3G 1A4 Montreal, Quebec, Canada

^b Department of Clinical Epidemiology, Universidad Autónoma de San Luis Potosí, 2405 Venustiano Carranza, 78290 San Luis Potosí, Mexico

^c Department of Orthopaedic Surgery, "Dr. Ignacio Morones Prieto" Central Hospital, 2395 Venustiano Carranza, 78290 San Luis Potosí, Mexico

ARTICLE INFO

Article history:

Received 16 August 2016

Available online 24 August 2017

Keywords:

Acetabular fracture

Functional outcome

Modified merle d'Aubigne-Postel score

Harris hip score

ABSTRACT

Introduction: The Modified Merle d'Aubigne-Postel Score and the Harris Hip Score are commonly used to assess the functional outcomes after acetabular fractures. A previous report showed that correlation between scores is good, that there is poor concordance among functional classes, and that the distribution of the scores is highly asymmetrical. Several issues were not addressed in this report, mainly that the data set was treated as transversal data without comparison of scores through time; therefore the objective of this article is to assess the degree of correlation and concordance between the Modified Merle d'Aubigne-Postel Score and the Harris Hip Score during the first 9 months after acetabular fracture treatment.

Methods: Both scores were recorded in a cohort of 23 previously healthy patients after 3, 6 and 9 months after fixation of acetabular fractures. Through a mixed-effects repeated measures model, we assessed differences between standardized scores. Pearson's interclass correlation coefficients for full scores and each of their domains, as well as agreement for clinical graduation classification was calculated.

Results: Between score correlation was 89%. We found no differences between scores at 3, 6 and 9 month follow-ups. Agreement between scores was 0.95, while agreement for clinical graduation classification was 0.67.

Discussion: Very short term correlation and concordance between scores is excellent, while concordance for clinical graduation classification is modest. We suggest the widespread use of the simpler score.

© 2017

1. Introduction

Acetabular fractures are relatively uncommon injuries, accounting for 0.3–6% of all fractures, and are associated to high energy traumas, mainly in the context of traffic accidents or falls from height.^{2,5} In order to assess the functional outcomes after these injuries, the Modified Merle d'Aubigne-Postel Score and the Harris Hip Score are commonly used. Both scores evaluate changes in hip function through the evaluation of 3 domains in the Modified Merle d'Aubigne-Postel Score (pain, ambulation and

range of movement (ROM)), and 4 in the Harris Hip Score (pain, function, absence of deformity and ROM).^{1,3}

Correlation between scores for patients assessed between 6 months and 10 years was found to be $R=0.81$ (Spearman's correlation coefficient) in one report.¹⁰ The authors noted that even though there was a high degree of correlation, concordance analysis with a kappa test was rather poor ($\kappa=0.45$), and that the distribution of the scores was highly asymmetrical, with values grouping towards the upper end of the scores. The authors concluded that a cautious approach should be used for comparing outcomes in smaller data sets, and that the score's design could be skewed, favouring the classification of patients into the higher ends of the score systems. To our opinion, several issues were not addressed in this report: first, the data set was treated as transversal data, although 661 observations of 450 patients were available. This did not allow for comparison of scores through time. Second, they started following patients after 6 months of

* Corresponding author at: Bone Engineering Labs of the Research Institute of the McGill University Health Centre, 1650 Cedar Ave, H3G 1A4 Montreal, Quebec, Canada.

E-mail address: jose.ramirezgarcialuna@mail.mcgill.ca (J.L. Ramirez-Garcialuna).

Table 1
Patient's clinical characteristics.

Variable	Value
Age (years)	33.91 SD 10.73
Sex	Male = 16 (69.57%) Female = 7 (30.43%)
Mechanism of injury	Traffic accident = 16 (69.57%) Run over injury = 4 (17.39%) Height fall = 3 (13.04%)
Type of fracture	Transverse = 10 (43.47%) Posterior wall = 6 (26.13%) Anterior column = 2 (8.69%) Posterior column = 2 (8.69%) Posterior column with posterior wall = 1 (4.34%) Transverse with posterior column = 1 (4.34%) Anterior column with posterior hemitransverse = 1 (4.34%)
Complications	None = 18 (73.92%) Associated sciatic nerve injury = 4 (17.39%) Erectile dysfunction = 2 (8.69%)

SD = standard deviation. Total of patients (n) = 23.

treatment, which may have allowed sufficient time for observations in the lower end of the scores to have been missed. Lastly, statistical analysis were based on the total numeric scores. This could represent a bias because scores have different maximum possible values. Spearman's correlation coefficient addresses this point, and comparison between data sets with different maximum possible values are not affected, but kappa tests are highly affected, and it may not be the best test of concordance for continuous data sets. In order to address this issues, and with the hypothesis that the best score is the simpler, the objective of this study was to compare the correlation between the Modified Merle d'Aubigne-Postel and Harris Hip Score at 3, 6 and 9 months in patients with acetabular fractures, and to compare their correlation by domains and for clinical graduation classification.

2. Material and methods

2.1. Patients

All previously healthy patients with an acetabular fracture who were treated in the "Dr. Ignacio Morones Prieto" Central Hospital of the Autonomous University of San Luis Potosí, Mexico, between 2009 and 2014 were recorded prospectively. The fractures were classified according to the Letournel and Judet classification⁸ and the patient's clinical characteristics are presented in Table 1. At each follow-up at 3, 6 and 9 months after fixation, clinical outcomes assessed through Modified Merle d'Aubigne-Postel Score³ and Harris Hip Score were recorded¹ (Table 2). The study was submitted and approved by the local Ethics Committee (registry 169872), and all patients agreed to participate in it.

2.2. Score analysis

The total numeric sum value for both scores were used in the present study. Clinical grades for the Modified Merle d'Aubigne-Postel Score were considered excellent at a cut-off value of 18, good at 15–17, fair at 12–14, and poor at 3–11 points⁴. For the Harris Hip Score, grades were considered excellent at 90–100, good at 80–89, fair at 70–79, and poor below 70 points¹⁰. In order to allow the comparison of two scores with different maximum values, we standardized them by dividing their values by the maximum possible value for each of them: 18 for the Modified Merle d'Aubigne-Postel Score, and 100 for the Harris Hip Score, thus obtaining a new numeric value representing a ratio of the full score.

First, we analysed through a lineal model the correlation between the total numeric score for both scores without adjusting for time, and without standardizing the scores. Afterwards, through a mixed-effects repeated measures lineal model we analysed the correlation between both standardized scores at 3, 6 and 9 months to obtain the correlation for same subject observations and assess for differences between scores. We calculated agreement between scores through the Pearson's interclass correlation coefficient (ICC) for average random raters (Shrout and Fleiss type 3.2⁹) for the full scores, and afterwards, for each of the three standardized domains shared by both scores (pain, function and ROM). Finally, to assess the correlation for clinical graduation classification, we performed a weighted kappa analysis for both full scores.

Statistical analysis was performed with R for Windows (version 3.0.2, R Core Team 2013) at 95% confidence level. Ethical approval was obtained from the Local Ethics Committee (registry 169862).

3. Results

From twenty seven patients admitted to the hospital with acetabular fractures during 2009 to 2014, twenty three were enrolled in the study. The reminder 4 patients were excluded because of previous comorbidities (1 diabetic, 2 with hypertension, and 1 with hypothyroidism). From the twenty three patients enrolled, 69 follow-ups were recorded, which sums a total of 138 observations (69 for the Merle d'Aubigne-Postel Score and 69 for the Harris Hip Score, Table 3). No patients died because of the injury or during follow-up.

Total score correlation, without accounting for time, was 89% ($R^2 = 0.89$, $p < 0.0001$). For each increase in 1 point in the Modified Merle d'Aubigne-Postel Score, there is a 5.65 (95%CI 4.91–6.39) point increase in the Harris Hip Score (Fig. 1). When accounting for time variation in the scores values, correlation for same subject observations was found to be 40.03%. We found no statistical significant differences between both standardized scores at 3 (0.53 ± 0.13 vs. 0.44 ± 0.17 , $p = 0.065$), 6 (0.70 ± 0.16 vs. 0.69 ± 0.17 , $p = 0.98$) and 9 months (0.81 ± 0.15 vs. 0.84 ± 0.16 , $p = 0.49$) (Fig. 2). Modified Merle d'Aubigne-Postel Score increase by month was 0.16 (95%CI 0.06–0.27) standardized points from month 3 to 6, and 0.11 (95%CI 0.008–0.21) from month 6 to 9, whereas Harris Hip Score increase was 0.25 (95%CI 0.14–0.36) standardized points from month 3 to 6, and 0.14 (95%CI 0.02–0.26) from month 6 to 9.

Table 2

Modified Merle d'Aubigne-Postel Score and Harris Hip Score.

		Scores
Modified Merle d'Aubigne-Postel Score (possible values: 3–18)		
Domain	Score	
Pain	No pain	6
	Slight or intermittent	5
	Pain after ambulation, but it disappears	4
	Moderately severe, permits ambulation	3
	Severe with ambulation	2
	Severe, prevents ambulation	1
Ambulation	Normal	6
	No cane, but slight limp	5
	Long distances with cane or crutch	4
	Limited, even with support	3
	Very limited	2
	Bedridden	1
Range of movement (%)	95–100	6
	80–94	5
	60–79	4
	40–59	3
	0–39	1
Harris Hip Score (possible values: 0–100)		
Pain	None or ignores it	44
	Slight, occasional, no compromise in activities	40
	Mild pain, no effect on average activities, rarely	30
	moderate pain with unusual activity, may take aspirin	
	Moderate pain, tolerable but makes concessions to pain. Some limitation of ordinary activity or work. May require occasional pain medicine stronger than	20
	Aspirin	
	Marked pain, serious limitations of activities	10
	Marked pain, serious limitations of activities	0
Function	A. Gait	
	1.- Limp	None 11
		Slight 8
		Moderate 5
		Severe 0
	2.- Support	None 11
		Cane for long walks 7
		Cane most of the time 5
		One crutch 3
		Two canes 2
		Two crutches 0
		Unable to walk 0
	B. Activities	
	1.- Stairs	Normally without using any railing 4
		Normally using a railing 2
		In any manner 1
		Unable to use stairs 0
	2.- Shoes and socks	With ease 4
		With difficulty 2
		Unable 0
	3.- Sitting	On ordinary chair for at least 1 h 5
		On a high chair for one-half hour 3
		Unable to sit comfortably in any chair 0
	4.- Distance	Unlimited 11
		Six blocks 8
		Two or three blocks 5
		Indoors only 2
		Bed and chair 0
Absence of deformity	5.- Able to enter public transportation	1
	4 points are given if the patient demonstrates:	4
	a. Less than 30° fixed flexion contracture	
	b. Less than 10° fixed adduction	
	c. Less than 10° fixed internal rotation in extension	
Range of movement	d. Limb-length discrepancy less than 3.2 cm	
	Else, the patient receives 0 points	
	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:	5
	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	

The overall numeric score is given by adding the domain scores. Clinical grades for the Modified Merle d'Aubigne-Postel Score: Excellent 18, Good 15–17, Fair 12–14, Poor 3–11 (4), and for the Harris Hip Score: Excellent 90–100, Good 80–89, Fair 70–79, Poor <70 (10).

Table 3

Patient scores.

Month	mMAPS total	mMAPS pain	mMAPS ambulation	MAP ROM	HHS total	HHS pain	HHS function	HHS deformity	HHS ROM
3	9.5 ± 2.3	3.3 ± 0.7	2.6 ± 0.7	3.5 ± 1.2	44.6 ± 16.6	26.0 ± 7.8	14.9 ± 8.1	0.3 ± 1.1	3.2 ± 1.1
6	12.5 ± 2.8	4.1 ± 0.8	4.0 ± 1.0	4.3 ± 1.1	69.4 ± 17.0	34.5 ± 7.4	29.2 ± 8.9	1.7 ± 2.0	3.9 ± 0.9
9	14.5 ± 2.6	5.0 ± 0.9	4.6 ± 0.9	4.8 ± 1.0	83.8 ± 16.3	39.4 ± 4.8	37.6 ± 10.1	2.4 ± 1.9	4.2 ± 0.8

mMAPS: Modified Merle d'Aubigne-Postel Score, ROM: Range of movement, HHS: Harris Hip Score. All values are presented as mean and standard deviation.

Overall Pearson's ICC between scores was found to be 0.95 (95% CI 0.92–0.97). Agreement by domains was 0.89 (95%CI 0.73–0.94) for pain, 0.92 (95%CI 0.86–0.95) for function, and 0.77 (95%CI 0.62–0.86) for ROM (Fig. 3). Agreement for clinical graduation classification was found to be weighted kappa = 0.67 (95%CI 0.58–0.76).

Overall score's values followed a normal distribution, as assessed by a Shapiro-Wilk normality test ($W = 0.983$, $p = 0.47$).

4. Discussion

We found no statistical differences between the Modified Merle d'Aubigne-Postel and the Harris Hip Score for assessing outcomes of patients with acetabular fractures at very short term. Overall correlation and agreement between scores was excellent, according to Landis and Koch criteria.⁷ This finding was maintained in the pain and function domain, but not in the ROM domain, where agreement was less than 0.80. Moreover, agreement for clinical graduation classification was found to be even lower. We believe that these findings could be due to the exclusion of the fourth dominion in the Harris Hip Score, absence of deformity criteria. Nevertheless, we believe that since both scores give mainly the same information, the Modified Merle d'Aubigne-Postel score should be used as routine because it is simpler, and that it is a misconception that the Harris Hip Scores “offers more information”.

Agreement for clinical graduation classification was found to be modest, although overall concordance was very good. The most likely explanation for this finding is that the cut-off values for the scores are not equal, and further studies should be performed to find the most appropriate value that best represents the patient's condition, and serves as the best prognostic marker.

In discordance with previous reports^{6,10}, we did not found a skewed distribution of our data. The most likely explanation for this finding is that since we included patient's records from very early phases after surgery, we can observe poor outcomes, and how they improve. The ceiling effect that they described is found after the sixth month, thus disproving the theory that the instruments are skewed in design (Fig. 1). We believe that since patients with acetabular fractures tend to be young and without comorbidities, the rapid appearance of the ceiling effect is due to a good clinical evolution.

Although the number of patient in our series is rather modest, we believe that we compensate this point with the inclusion of time as a covariate. As far as we know, this is the first report to compare functional outcome scores in patients with acetabular fractures through time, and not as transversal data.

5. Conclusions

The Modified Merle d'Aubigne-Postel score seems to be very similar to the Harris Hip Score for graduation of patient's outcomes through the first nine months after an acetabular fracture. Overall correlation and agreement for the pain and function domains are excellent, whereas agreement for ROM and clinical graduation classification are modest. Since we used ICC for average random raters, conclusions can be extended for any rater and any patient.

Conflict of interest and funding

All authors declare: no financial relationships with any organizations that might have an interest in the submitted work; no other relationships or activities that could appear to have influenced the submitted work.

Fig. 1. Correlation between scores.

Scatter plot between the total numeric score of the modified Merle d'Aubigne-Postel Score (mMAPS) and the Harris Hip Score (HHS). Correlation was found to be $R^2 = 0.89$, $p < 0.0001$. A ceiling effect can be observed after the 6th month.

Fig. 2. Comparison of standardized scores through time.

Mean standardized scores and their 95% confidence intervals are shown for the modified Merle d'Aubigne-Postel Score (mMAPS) and the Harris Hip Score (HHS). No statistical significant differences are observed.

Fig. 3. Bland-Altman plots for agreement between scores.

Concordance between standardized scores are presented. Pearson's ICC between scores was 0.95, 0.89 for pain domain, 0.92 for function domain, and 0.77 for ROM domain.

References

- Harris WH. 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;51:737–755.
- Laird A, Keating JF. Acetabular fractures: a 16-year prospective epidemiological study. *J Bone Joint Surg Br.* 2005;87:969–973.
- Matta JM, Mehne DK, Roffi R. Fractures of the acetabulum: early results of a prospective study. *Clin Orthop.* 1986;1:241–250.
- Marchetti P, Binazzi R, Vaccari V, et al. Long-term results with cementless fiteck (or fitmore) cups. *J Arthroplasty.* 2005;20:730–737.
- Mauffrey C, Hao J, Cuellar DO, et al. The epidemiology and injury patterns of acetabular fractures: are the USA and China comparable? *Clin Orthop.* 2014;472:3332–3337.
- Moed BR, Yu PH, Gruson KI. Functional outcomes of acetabular fractures. *J Bone Joint Surg Am.* 2003;85-A:1879–1883.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–174.
- Letournel E, Judet R, Elson R. *Fractures of the Acetabulum*. 2nd ed. Germany : Berlin; New York: Springer-Verlag; 1993 [xxiii, 733 p.: ill].
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86:420–428.
- Øvre S, Sandvik L, Madsen JE, Røise O. Comparison of distribution, agreement and correlation between the original and modified Merle d'Aubignè-Postel Score and the Harris Hip Score after acetabular fracture treatment: moderate agreement, high ceiling effect and excellent correlation in 450 patients. *Acta Orthop.* 2005;76:796–802.