**Surprisingly poor correlation among pain scores prior to total hip arthroplasty.**

K Ions\* FRCS1

L Jonker PhD2

1 Department of Orthopaedic Surgery, North Cumbria University Hospitals NHS Trust, Newtown Road, Carlisle, CA2 7HY, UK.

2 Senior research fellow, Faculty of Health, Medical Sciences and Social Care, University of Cumbria, Fusehill Street, Carlisle, CA1 2HH, UK

**Running title: pain scores prior to total hip arthroplasty.**

**Please address all correspondence to:**

\*Keith Ions, Consultant Orthopaedic Surgeon, Department of Orthopaedic Surgery, North Cumbria University Hospitals NHS Trust, Newtown Road, Carlisle, CA2 7HY, UK. G.Ions@ncumbria-acute.nhs.uk , Tel. 01228 523444, Fax 01228 814689.

**Conflict of interest statement**

 No funds were received in support of this study.

**ABSTRACT**

Pain is the main indication for total hip arthroplasty but is a subjective experience and can therefore be difficult to assess. Using completed forms from six hundred consecutive patients who were to undergo a total hip arthroplasty, we investigated the level of correlation between three widely used pain rating scales. These are: visual analogue pain scale, Merle d’Aubigne pain score and the first question on the Oxford Hip Score. We demonstrate here that there is a surprisingly poor correlation between the scores obtained from the three rating scales, albeit statistically significant. The use of more than one of the above pain scores is indicated due to the slightly different ways in which they ask the patients about their pain experience. This undoubtedly contributes to the modest correlation levels observed. Drawing extensive conclusions from a single pain rating scale may have significant clinical implications.

[143 words]

**Key words:** total hip arthroplasty, pain, pain score, rating scale, correlation.

**Introduction**

Various scoring systems have been used to assess patients before and after total hip replacement surgery [1 - 3]. Although additional variables are measured in some hip scores, the core variables measured are pain, movement and function. The contribution of the variable ‘pain’ to the total score is often larger than that of movement and function. Clinically, the most important feature deciding who requires surgery is usually pain not adequately controlled by analgesia. Bryant and colleagues [4] examined thirteen methods of hip scoring, including the Merle d’Aubigne score, and from their statistical analysis they concluded that only the aforementioned three variables need to be recorded. One of these was pain, which correlated weakly with the other variables.

The Oxford Hip Score (OHS) was developed for use in patients who require total hip replacement [5]. In contrast to the outcomes observed by Bryant, when validating the OHS against other established clinical assessment tools, including the Charnley, SF36 and Arthritis Impact Measurement Scales (AIMS) clinical assessments, it was shown that pain was actually the element with the highest correlation between the different assessment tools. From this, it appears there is a discrepancy in terms of the value of a pain score obtained with different clinical assessment tools.

Despite the availability of numerous clinical assessment tools for total hip arthroplasty, incorporating pain scores, and stand alone pain rating scales, there does not seem to be a consensus on which pain rating scale, or combination of pain rating scales, should be applied in clinical practice, particularly pre-operatively, or indeed how well they correlate. We evaluated how well three pain rating scales correlate when completed by the same patient. In addition, we have also compared these pain rating scales with one of the most commonly applied clinical assessment questionnaire for total hip arthroplasty, the OHS.

**Method**

In our department, a database of patients undergoing total hip replacement has been kept since June 2005. In total, pain rating scores for 602 patients completed prior to total hip arthroplasty were used in our analysis. It is standard practice in our clinics to ask patients to complete a visual analogue pain rating scale, a Charnley modified Merle d’Aubigne pain score (MDA) and an Oxford Hip Score (OHS total). The first question on the Oxford Hip Score is about pain (OHS pain). All patients were asked to complete the questionnaires in the order given above. We used the OHS pain score plus the other two pain scores to provide three measures of pain for correlation analysis. In addition, we also looked at the relationship of these scores to the OHS total score obtained for all patients. Statistical analysis, non-parametric correlation with the Spearman test, was performed using SPSS version 17 (SPSS Inc., Chicago, USA).

**Results**

Of the three pain rating scales investigated, OHS pain and MDA are ordinal scales. The visual analogue pain scale can be classed as continuous. Descriptive statistical analysis on the distribution of the pain scores revealed a Gaussian distribution for the visual analogue and Charnley modified MDA score and total Oxford hip score (data not shown). However, the pain element of the Oxford hip score is negatively skewed with virtually all scores being the maximum outcome (patients usually experience ‘severe’ pain from their hip). The basic measures of central tendency and dispersion are given in Table 1.

[Table 1 to be inserted here]

A subsequent calculation of the correlation coefficients between the various pain scores, and between each pain score and the total OHS shows that the correlations range from 0.42 (Visual analogue score vs total OHS score) to 0.56 (Oxford score pain element vs total OHS score).

 [Table 2 to be inserted here]

**Discussion**

It is recognised that pain is very difficult to measure [6]. In contrast, other key symptoms and signs of disorders affecting the hip, movement and function, can often be easily measured – e.g. movement range of hip flexion. The correlations found between the three pain rating scales of around 0.5 are surprisingly low considering that the scores are attempting to measure the same item, namely pain. In the context of looking at the correlation between two separate variables, 0.5 may be considered on the border between medium/modest (0.30-0.49) and large/strong (0.50-1.00) [7]. However, 0.5 in this context can perhaps be considered a poor correlation, despite the fact that ordinal data tends to result in lower correlation coefficients due to fewer measuring points. Of interest is the inclusion of the total OHS and its comparison to the pain-specific rating scales. One would expect the pain rating scales to correlate closer to each other than the total OHS, even if the pain element is obviously a contributor to the total OHS score.

There are a few factors that may have contributed to the relatively modest correlations observed here. Firstly, as mentioned above, two of the three pain rating scales use ordinal data (OHS pain and MDA). Secondly, the three pain scales use quite widely differing language in their questions about pain. For example, the pain element of the Oxford hip score states ‘severe’ as the maximum score. On the other hand, the MDA score states ‘intense and permanent’ as the maximum score followed by ‘severe’ pain, which has been sub-divided into severe ‘whilst being active’ and severe ‘even at night’. The third observation is that there is no temporal aspect to the visual analogue score and MDA score whereas the OHS (including Oxford pain element) asks the patient to sum up the pain experienced over the past four weeks. A fourth and final note is whether completing a visual pain rating scale leads to different outcomes than completing a numerical pain rating scale. All these variations also impact on the validity to use correlation coefficient measurement as a means to compare scores on different pain rating scales. However, correlation analysis was also used by Bryant and colleagues and also by Dawson and colleagues when developing the Oxford hip score [4, 5].

In this study we consciously did not concentrate on the change in pain experienced by patients before and after total hip arthroplasty, because the aim here was to determine whether different pain rating scales and questionnaires are equally appropriate and comparable in identifying those patients that require total hip arthroplasty in the first place. The decision to operate is made by the surgeon based on all the information obtained at the consultation, including details from the history the examination and radiography diagnostics. However pain not relieved by adequate analgesia remains a major factor in the decision to operate in most cases.

Our data shows that depending on the pain rating scale used, the ultimate clinical decision could well be different if the assessment was based entirely on a score rather than a full clinical assessment. Since we are not able to produce measurements of pain which correlate strongly with each other, and pain is one of the most important elements in hip scores, the use of scores alone to list patients for surgery is problematic. It may be advisable to use more than one form of pain score in the assessment of patients considered for such surgery [6]. Reliance on one score, especially if numerical scoring were to be used as a tool to decide which patients should have total hip arthroplasty, could seriously disadvantage some patients who may score much lower on one pain score than another. Experienced surgeons are likely to make an assessment of a patient’s pain level from the history, and perhaps an interesting study would be for the surgeon to do an analogue pain score on the patient’s behalf, blind to the patients’ scores and look at the correlation. The processing of information by the surgeon may deal with the discrepancies seen in pain scores. In times when scoring is being taken from its natural use as a tool to assess and compare results, to being touted as a means to decide who needs operation, the role of the surgeons assessment needs to be clarified to avoid potential of scoring causing some patients to lose out.

REFERENCES

1. Harris WH. (1969) 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. 51: 737-55.

2. Ritter MA, Fechtman RW, Keating EM, Faris PM. (1990) The use of a hip score for evaluation of the results of total hip arthroplasty. J Arthroplasty 5: 187-9

3. Johanson NA, Charlson ME, Szatrowski TP, Ranawat CS. (1992) A self-administered hip-rating questionnaire for the assessment of outcome after total hip replacement. J Bone Joint surg [Am] 74-A: 587-97.

4. Bryant MJ, Kernohan WG, Nixon JR, Mollan RAB. (1993) A statistical analysis of hip scores. J Bone Joint Surg [Br] 75-B: 705-9.

5. Dawson J, Fitzpatrick R, Carr A, Murray D. (1996) Questionnaire on the perceptions of patients about total hip replacement. J Bone Joint Surg [Br] 78: 185-90

6. Bradley LA. (1993) Pain measurement in arthritis. Arthritis Care Res. 6: 178-86.

7. Kozak M. (2009) What is strong correlation? Teaching Statistics. 31: 85-86

**Table 1**. Overview of measures of central tendency and dispersion for each pain scale.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Type of pain rating scale** | **Mean**  | **Median** | **SD**  | **Observed (min-max)** |
| Visual analogue score | 7.3  | 7.6  | 1.6  | 1.4 – 10  |
| MDA score - pain | 2  | 2  | 1  | 5 – 0  |
| Oxford score - pain | 5  | 5  | 1  | 2 – 5  |
| Oxford score - total | 44  | 45  | 6  | 16 – 60  |

Scale ranges for the different measuring tools are: visual analogue score, 0 to 10; MDA score, 6 to 0; Oxford score – pain element, 1 to 5; Oxford score – total, 12 to 60. SD, standard deviation.

**Table 2**. Correlation levels between different pain rating scales.

|  |  |
| --- | --- |
| **Comparison test** | **Correlation coefficient\*** |
| Visual analogue score vs MDA score | * 0.47
 |
| Visual analogue score vs Oxford score total |  0.42 |
| Visual analogue score vs Oxford score pain |  0.53 |
| MDA score vs Oxford score total | * 0.48
 |
| Oxford score pain vs MDA |  0.53 |
| Oxford score pain vs Oxford score total  |  0.56 |

 \*All correlation outcomes were statistically significant with p<0.01.