BJR

ARTHROPLASTY

Responsiveness and ceiling effects of the Forgotten Joint Score-12 following total hip arthroplasty

D. F. Hamilton,
J. M. Giesinger,
D. J. MacDonald,
A. H. R. W. Simpson,
C. R. Howie,
K. Giesinger

University of Edinburgh, Edinburgh, United Kingdom

■ D. F. Hamilton, PhD, BSc (Hons), BSc (Hons), Research Fellow, Department of Orthopaedics, University of Edinburgh, FU413, Chancellor's Building, 49 Little France Crescent, Edinburgh, EH164SB, UK.
■ J. M. Giesinger, PhD, Department of Psychiatry and Psychotherapy, Research Fellow, Department of Psychiatry and Psychotherapy, Innsbruck Medical University, Anichstrasse 35, 6020 Innsbruck, Austria.
■ D. J. MacDonald, MIQA, CQP, Research Associate, Department of Orthopaedics, University of Edinburgh, FU413, Chancellor's Building, 49 Little France Crescent, Edinburgh, EH164SB, UK.
■ A. H. R. W. Simpson, FRCS, MA(Cantab), DM(Oxon), Professor of Trauma and Orthopaedics, University of Edinburgh, FU413, Chancellor's Building, 49 Little France Crescent, Edinburgh, EH164SB, UK.
■ C. R. Howie, FRCS, Professor of Orthopaedics, Department of Orthopaedics, University of Edinburgh, FU413, Chancellor's Building, 49 Little France Crescent, Edinburgh, EH164SB, UK.
■ K. Giesinger, MD, MSc, Orthopaedic Surgeon, Department of Orthopaedics, Kantonsspital St. Gallen, Rorschacherstrasse 95, 9007 St. Gallen, Switzerland.

Correspondence should be sent to Dr D. F. Hamilton; e-mail: d.f.hamilton@ed.ac.uk

doi:10.1302/2046-3758.53.2000480

Bone Joint Res 2016;5:87–91.
Received: 29 June 2015;
Accepted: 15 February 2016

Objectives

To assess the responsiveness and ceiling/floor effects of the Forgotten Joint Score -12 and to compare these with that of the more widely used Oxford Hip Score (OHS) in patients six and 12 months after primary total hip arthroplasty.

Methods

We prospectively collected data at six and 12 months following total hip arthroplasty from 193 patients undergoing surgery at a single centre. Ceiling effects are outlined with frequencies for patients obtaining the lowest or highest possible score. Change over time from six months to 12 months post-surgery is reported as effect size (Cohen's d).

Results

The mean OHS improved from 40.3 (SD 7.9) at six months to 41.9 (SD 7.2) at 12 months. The mean FJS-12 improved from 56.8 (SD 30.1) at six months to 62.1 (SD 29.0) at 12 months. At six months, 15.5% of patients reached the best possible score (48 points) on the OHS and 8.3% obtained the best score (100 points) on the FJS-12. At 12 months, this percentage increased to 20.8% for the OHS and to 10.4% for the FJS-12. In terms of the effect size (Cohen's d), the change was d = 0.10 for the OHS and d = 0.17 for the FJS-12.

Conclusions

The FJS-12 is more responsive to change between six and 12 months following total hip arthroplasty than is the OHS, with the measured ceiling effect for the OHS twice that of the FJS-12. The difference in effect size of change results in substantial differences in required sample size if aiming to detect change between these two time points. This has important implications for powering clinical trials with patient-reported measures as the primary outcome.

Cite this article: *Bone Joint Res* 2016;5:87–91.

Keywords: hip arthroplasty; outcomes; PROM: forgotten joint score

Article focus

- The use of outcome measures to qualify success following total hip arthroplasty; specifically to assess the responsiveness of the Forgotten Joint Score – 12 in a United Kingdom population.

Key messages

- The FJS-12 is more responsive to change between six and 12 months after total hip arthroplasty and demonstrates half the ceiling effects of the Oxford Hip Score.
- Use of a more sensitive score has material implications in detecting differences in patient outcome over time.
- Use of an outcome measure with greater measurement range and precision allows for more specific studies with smaller numbers of participants, which has significant benefits for researchers and funders, allowing for quicker and cheaper studies.

Strengths and limitations

- This study is the first specific evaluation of the methodological qualities of the FJS-12 in a United Kingdom or European population of hip arthroplasty patients.
- The FJS-12 was described as a post-operative outcome score, however the use of pre-operative data, would offer additional perspective in terms of assessing change in score following surgery.

Introduction

Patient-reported outcome measures (PROMs) are a key parameter for assessing outcome following total hip arthroplasty. PROMs are widely used to power clinical studies and increasingly complement implant survivorship in national joint registries, providing a measure of a patient's pain and physical function.[1,2]

There is a plethora of potential patient outcome tools that can be employed, which, though similar, offer a unique perspective on patient outcome. There is no consensus as to which scores 'should' be used, and researchers typically revert to the most commonly reported scores such as (for hip arthroplasty) the Oxford Hip Score (OHS)[3] or Harris Hip Score[4]

PROM questionnaires differ substantially in their ability to capture change following joint replacement surgery.[5,6] The sensitivity of a score to detect change is critical in determining the level of function a patient has achieved. Ceiling effects in a scoring system can hide differences when patients already record the maximum possible score and cannot improve on that score. For example, in a score that simply assesses whether patients can walk for more than 30 minutes without pain, a 75-year-old patient able to walk for just 30 minutes at a normal pace would have the same score as a 45-year-old patient who has returned to running marathons.

When designing clinical outcome studies, the responsiveness of a score is of key importance in determining which score is used to power the study, as this materially affects the sample size required. Responsive scores are required as the symptom burden and range of functional impairments define whether ceiling effects will reduce the discriminatory power of the employed PROM. In terms of assessing the outcome of joint arthroplasty, a measure should have a measurement range that is responsive across the patient's surgical pathway, including pre-operative and longitudinal post-operative outcomes.

A new joint-specific score, the Forgotten Joint Score-12 (FJS-12),[7] was introduced in 2012 to asses a novel construct, 'joint awareness'. The authors of this score suggested that the new construct would be more responsive to higher level functional outcomes after joint arthroplasty, because being able to 'forget' the joint during daily life is perhaps the ultimate goal following joint replacement surgery.

The objective of this study was to assess the responsiveness and ceiling effects of the Forgotten Joint Score-12 and compare these measures with that of the more widely used Oxford Hip Score in patients six and 12 months following primary total hip arthroplasty.

Patients and Methods

Sample. Patients undergoing total hip arthroplasty over a six-month period (January 2013 to June 2013) at a single large orthopaedic teaching hospital were assessed six and 12 months post-surgery using the OHS and the FJS-12. The study centre is the only hospital receiving adult referrals for a predominantly urban population of approximately 850 000. Data had been collected through informed consent for inclusion in a departmental database for which regional ethical approval had been obtained (ref 11/AL/0079). Procedures were carried out by multiple consultant orthopaedic surgeons and their supervised trainees. All data was collected independently from the clinical team by the arthroplasty outcomes research unit of the associated university. All patients completed pre-operative outcome questionnaires in clinic and by postal follow-up questionnaires at six and 12 months post-operatively. Demographic data including age, gender and comorbidities, were reported by the patient as part of the survey.

Assessment instruments. The Forgotten Joint Score-12 is a recently published patient-reported outcome scale designed to assess joint awareness in hips and knees during various activities of daily living. It uses a five-point Likert response format, consisting of 12 equally weighted questions with the raw score transformed to range from zero to 100 points. High scores indicate good outcome, i.e. a high degree of being able to forget about the affected joint in daily life. In previous studies the score has shown good reliability and convergent validity,[7,8] performed well in known-group comparisons, and was found to be sensitive to change over time.[6]

The Oxford Hip Score consists of 12 questions relating to the patient's perceived pain and functional ability, answered on a Likert scale with values from zero to four. The score ranges from zero to 48, with the overall score calculated from the responses to the 12 questions. A score of zero is the worst possible outcome, suggesting severe symptoms and dysfunction, while 48 is the best possible outcome. It is widely used and has been shown to have good reliability, validity and responsiveness to clinical change.[5,9]

Statistical analysis. Descriptive statistics are presented as frequencies, means, standard deviations (SD) and ranges. Ceiling effects are described as the frequencies for patients obtaining the highest possible score on a PROM. Change over time, from six months to 12 months post-surgery, are presented as effect size (Cohen's d). Effect sizes were assessed for the whole cohort and in age, gender, and comorbid case-mix subgroups. Statistical analysis was conducted with SPSS 21.0 and power analysis with G* Power 3.1.[10]

Results

Patient characteristics. Within the study period we collected OHS and FJS-12 data from 193 patients. Mean age was 67.6 years (SD 10.5) and 116 (60.1%) were female. Body mass index at study inclusion was 27.7 on average (SD 5.2). The most frequent self-reported comorbid

Table I. Patient characteristics (n = 193).

Mean age (SD); range	**67.6 (10.5); 28 to 91**
Gender **(%)** women/men	116 *(60.1)*/77 *(39.9)*
Side (%): left/right	*57.6/42.4*
Mean body mass index (SD); range	27.7 (5.2) 18.2 to 60.6
Comorbidity (%): back pain/high blood pressure/depression/heart disease/ diabetes/cancer/liver disease/stomach ulcer/anaemia/kidney disease/liver disease	94 *(56.0)*/72 *(41.1)*/33 *(20.6)*/20 *(11.8)*/16 *(9.6)*/12 *(7.5)*/11 *(6.8)*/10 *(6.3)*/7 *(4.4)*/2*(1.2)*/1 *(0.6)*

SD, standard deviation

Table II. Descriptive statistics for the Oxford Hip Score and the Forgotten Joint Score–12 at six and 12 months post-surgery.

	6 months	**12 months**	
Mean Oxford Hip Score (SD); median; range	40.3 (7.9); 42; 10 to 48	41.1 (7.2); 43; 17 to 48	Effect size d = 0.10 Sample size* n = 787
Mean Forgotten Joint Score–12 (SD); median; range	56.8 (30.1); 56; 0 to 100	62.1 (29.0); 65; 0 to 100	Effect size d = 0.17 Sample size* n = 262

*Sample size required to detect a mean change between six and 12 months (*t*-test for dependent samples, power = 0.80, alpha = 0.05, two-sided);
SD, Standard deviation

conditions were back pain (56.0%), high blood pressure (41.1%), depression (20.6%) and heart disease (11.8%). Of these patients, 19.9% reported no comorbid conditions, 37.6% one comorbidity, and 42.5% two or more comorbidities (Table I).

Sensitivity to change over time. The mean OHS improved from 40.3 (SD 7.9) at six months to 41.9 (SD 7.2) at 12 months. Mean FJS-12 improved from 56.8 (SD 30.1) at six months to 62.1 (SD 29.0) at 12 months (Table II). At six months, 15.5% of the patients reached the best possible score (48 points) on the OHS and 8.3% obtained the best score (100 points) on the FJS-12. At 12 months, this percentage increased to 20.8% for the OHS and to 10.4% for the FJS-12 (Figs 1 and 2).

In terms of the effect size (Cohen's d), the change was d = 0.10 for the OHS and d = 0.17 for the FJS-12. In a power analysis for the comparison of two time points (*t*-test for dependent sample, power = 0.80, alpha = 0.05, two-sided), the required sample size to detect a difference between six and 12 months is 787 patients for the OHS and 262 patients for the FJS-12.

Subgroup analysis. We found the OHS to have the same change over time in men and women (Cohen's d = 0.10 in both groups). The FJS-12 was more sensitive to change in men (d = 0.22) than in women (d = 0.14). The FJS-12 showed higher effect sizes than the OHS (0.26 vs 0.19) in patients aged below 70 years. In those patients aged above 70 years there was only very little improvement found in this follow-up period with an effect size of d = 0.01 for the OHS and d = 0.07 for the FJS-12.

A large proportion of patients reported a single comorbidity (high blood pressure), which was unlikely to influence patient-reported pain and function following hip replacement. We therefore dichotomised our dataset based on patients having no/one comorbidity or multiple comorbidities. This produced similarly sized groups, which offered the best estimates for effect sizes. We found that in patients with no/one comorbid condition, the

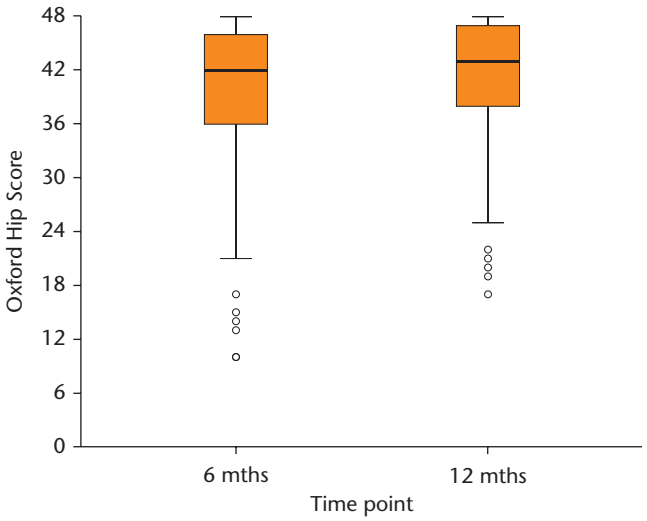

Fig. 1

Boxplot of the Oxford Hip Score at six and 12 months post-operatively

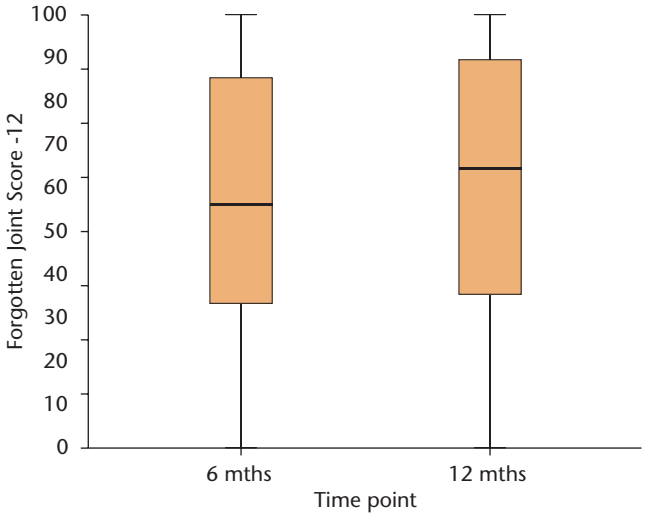

Fig. 2

Boxplot of the Forgotten Joint Score-12 at six and 12 months post-operatively

FJS-12 had higher sensitivity to change than the OHS (d = 0.25 vs d = 0.11), whereas in those patients with two or more comorbid conditions the OHS showed a similar change (d = 0.12 vs d = 0.10).

Discussion

These findings suggest that the FJS-12 is more responsive to change between six and 12 months after total hip arthroplasty than is the OHS. The measured ceiling effect for the OHS was twice that of the FJS-12.

A ceiling effect means that several patients score the highest possible score achievable with the employed test. A ceiling effect occurs when the test items are not challenging enough for a group of individuals because the test has a limited number of difficult items or even an inappropriate item selection.[11] This will lead to a shortcoming in the discriminative ability of the test to detect clinically relevant changes so that if the patient continues to improve, the test does not capture that improvement. A ceiling effect of 15% is considered the maximum acceptable cut-off value.[12]

We found minimal difference in patient post-operative outcome as measured with the OHS between six and 12 months post-hip arthroplasty, however, the FJS-12 reported a larger change between six and 12 months, perhaps more accurately reflecting the changing perception of the patient as to the impact of the THA on their quality of life. The subgroup comparisons further highlight this difference, where the FJS-12 offered greater sensitivity to change than did the OHS in younger patients (under 70), healthier patients and males, i.e. those most likely to achieve high scores.[13,14]

Outcome tools to assess joint arthroplasty first appeared in the late 1960s. When the Harris Hip Score was designed (in 1969), its content validity may well have been sufficient for the surgical case-mix of that time. However, more than 40 years later primary hip arthroplasty patients and the lifestyle of that age group has changed considerably. Wamper et al[15] found a relevant ceiling effect after hip arthroplasty in 31 out of 59 studies in their meta-analysis of studies reporting Harris Hip Score data. Similarly, the OHS is an ageing tool developed in the 1990s. It has been well validated, but is increasingly understood to suffer from ceiling effects.[9,12] A major difference between the FJS-12 and the OHS is the population for which they were designed. The OHS was developed in a pre-operative population[3] and reflects the symptom state at this time. The FJS-12, in contrast, was designed in a post-operative population and reflects the symptom state following surgery. This, in principle, should allow more measurement variation for higher level function and display a lesser ceiling effect, as is seen in Figures 1 and 2.

The Forgotten Joint Score was formally introduced in 2012[7] and various groups have described the measurement properties of the FJS-12 in knee arthroplasty;[6,8] however, there is a lack of literature as to the score measurement properties following total hip arthroplasty. The original paper described similar ceiling and floor effects to the current data in the United Kingdom (9.3 and 3.3 %, respectively) in a Swiss population, although at a longer mean follow-up period of 3 years. Matsumoto et al[16] recently validated a translated version of the FJS-12 in a Japanese hip arthroplasty population of 108 patients. These authors reported a single post-operative time point analysis comparing the FJS-12 to WOMAC and JEHQ scores at a mean follow-up of 30 months (with a wide follow-up time point range of one to 180 months). The authors reported a lower ceiling effect in the FJS-12 compared with the other measures, but did not report the percentage of patients achieving the top score, making direct comparisons difficult. Homma et al[17] included the FJS-12 as an outcome metric in a study evaluating the effect of the surgical approach on lateral femoral cutaneous nerve injury rates, again in a Japanese population, of 122 patients. Measurement properties of the score were not evaluated, but the outcome scores presented for patients (post-operative FJS-12 of 64.3) are similar to those which we report in this United Kingdom population (62.1) and to those which Behrend et al[7] reported in the original Swiss population (59.8).

The strengths of this study include the first presentation of linked post-operative FJS-12 outcome data at six and 12 months following total hip arthroplasty. The current study is also the first specific evaluation of the methodological qualities of this scoring system in a United Kingdom or European population of hip arthroplasty patients (aside from the original Swiss population). Limitations of this work include the restricted time points available for analysis. It would be useful to identify responsiveness over a longer post-operative timeframe, though 12 months is the typical outcome timeframe reported in the orthopaedic literature. In addition, the fact that the outcome scores presented herein are similar to those reported in differing populations at three years suggests that there is little further change in symptomology, although detailed investigation is required to confirm this. Furthermore, the FJS-12 was described as a post-operative outcome score and can be contrasted against healthy control population scores to offer context. Use of pre-operative data, however, would offer additional perspective in terms of assessing change in score following surgery. We suggest this as an avenue for future research.

Use of a more sensitive score that is not restricted by ceiling effects has material implications, both in detecting differences in patient outcome/function over time, and in the numbers of patients needed in a study to detect a difference in outcome between groups. The results we report herein suggest that around half the number of participants would be required to power a superiority design randomised trial when using the FJS-12 compared with

the OHS. There is seemingly little difference in the effect sizes we report (0.10 and 0.17) but this translates as very different sample size requirements in terms of numbers of participants; this is because the association between effect size and sample size is essentially exponential. Use of an outcome measure with greater measurement range and precision allows for more specific studies with smaller numbers of participants, which has significant benefits for researchers and funders, allowing for quicker and cheaper studies.

Our findings suggest that the FJS-12 is more responsive to change between six and 12 months after total hip arthroplasty than is the OHS. The measured ceiling effect for the OHS was twice that of the FJS-12. The difference in effect size of change results in substantial differences in required sample size if aiming to detect change between these two time points. This has important implications for powering clinical trials with patient reported outcomes as the primary outcome.

References

1. **Baker PN, van der Meulen JH, Lewsey J, Gregg PJ, National Joint Registry for England and Wales.** The role of pain and function in determining patient satisfaction after total knee replacement. Data from the National Joint Registry for England and Wales. *J Bone Joint Surg [Br]* 2007;89-B:893-900.
2. **Hamilton DF, Lane JV, Gaston P, et al.** What determines patient satisfaction with surgery? A prospective cohort study of 4709 patients following total joint replacement. *BMJ Open* 2013;3:pii: e002525.
3. **Dawson J, Fitzpatrick R, Carr A, Murray D.** Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg [Br]* 1996;78-B:185-190.
4. **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-A:737-755.
5. **Nilsdotter A, Bremander A.** Measures of hip function and symptoms: Harris Hip Score (HHS), Hip Disability and Osteoarthritis Outcome Score (HOOS), Oxford Hip Score (OHS), Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH), and American Academy of Orthopedic Surgeons (AAOS) Hip and Knee Questionnaire. *Arthritis Care Res (Hoboken)* 2011;63(Suppl 11):S200-207.
6. **Giesinger K, Hamilton DF, Jost B, Holzner B, Giesinger JM.** Comparative responsiveness of outcome measures for total knee arthroplasty. *Osteoarthritis Cartilage* 2014;22:184-189.
7. **Behrend H, Giesinger K, Giesinger JM, Kuster MS.** The "Forgotten Joint" as the Ultimate Goal in Joint Arthroplasty Validation of a New Patient-Reported Outcome Measure. *J Arthroplasty* 2012;27:430-436.
8. **Thompson SM, Salmon LJ, Webb JM, Pinczewski LA, Roe JP.** Construct Validity and Test Re-Test Reliability of the Forgotten Joint Score. *J Arthroplasty* 2015;30:1902-1905.
9. **Ostendorf M, van Stel HF, Buskens E, et al.** Patient-reported outcome in total hip replacement. A comparison of five instruments of health status. *J Bone Joint Surg [Br]* 2004;86-B:801-808.
10. **Faul F, Erdfelder E, Buchner A, Lang AG.** Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods* 2009;41:1149-1160.
11. **McHorney CA, Tarlov AR.** Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995;4:293-307.
12. **Terwee CB, Bot SD, de Boer MR, et al.** Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34-42.
13. **The NJR Editorial Board.** National Joint Registry. *National Joint Registry for England and Wales: 10th Annual Report*, 2013. http://www.njrcentre.org.uk/njrcentre/default.aspx (date last accessed 05 February 2016).[[bibmisc]]
14. **Hamilton DF, Giesinger JM, Patton JT, et al.** Making the Oxford Scores clinically meaningful through normative scoring. *Bone Joint Res* 2015;4:137-144.
15. **Wamper KE, Sierevelt IN, Poolman RW, Bhandari M, Haverkamp D.** The Harris hip score: Do ceiling effects limit its usefulness in orthopedics? *Acta Orthop* 2010;81:703-707.
16. **Matsumoto M, Baba T, Homma Y, et al.** Validation study of the Forgotten Joint Score-12 as a universal patient-reported outcome measure. *Eur J Orthop Surg Traumatol* 2015;25:1141-1145.
17. **Homma Y, Baba T, Sano K, et al.** Lateral femoral cutaneous nerve injury with the direct anterior approach for total hip arthroplasty. *Int Orthop* 2015. (Epub ahead of print)

Funding Statement:
- A. H. R. W. Simpson reports funding received from Styker which is related to this article. He reports further funding from ARUK, ORUK, OTCF, EPSRC and RCSEd, none of which is related to this article.
- C. R. Howie reports funding received from Styker/Biomet which is related to this article.
- D. F. Hamilton reports funding received from ARUK and Stryker which is not related to this article.
- J. M. Giesinger and K. Giesinger are copyright holders of the Forgotten Joint Score - 12. Royalties are payble for commercial use of the questionnaire - this is not related to this article.

Author Contribution:
- D. F. Hamilton, Study concept, data analysis and interpretation, writing of the manuscript, approval of the final manuscript.
- J. M. Giesinger, Study concept, data analysis and interpretation, writing of the manuscript, approval of the final manuscript.
- D. J. MacDonald, Study concept, data collection and analysis, editing and approval of the final manuscript.
- A. H. R. W. Simpson, Study concept, interpretation of data, editing of and approval of the final manuscript.
- C. R. Howie, Study concept, interpretation of data, editing and approval of the final manuscript.
- K. Giesinger, Study concept, data analysis and interpretation, writing of the manuscript, approval of the final manuscript.

ICMJE conflict of interest:
- C. R. Howie is the Past President of the British Orthopaedic Association.