The Journal of Arthroplasty 29 (2014) 491–494

Contents lists available at ScienceDirect

The Journal of Arthroplasty

journal homepage: www.arthroplastyjournal.org

Mail Versus Telephone Administration of the Oxford Knee and Hip Scores

Marena Abdel Messih, MBBS Student [a], J.M. Naylor, PhD, BAppSc (Phty) [b,c,d], J. Descallar, BSc, MBiostat [e,f], A. Manickam, MBBS Student [a], R. Mittal, BSc(Med), MBBS, MMed(Clin Epi), MS [b,c,d], I.A. Harris, MBBS, MMed(Clin Epi), PhD, FRACS(Orth), FAOrthA [b,c,d]

[a] University of New South Wales, Kensington, Sydney, Australia
[b] South Western Sydney Clinical School, University of New South Wales, Liverpool, NSW, Australia
[c] Liverpool Hospital Orthopaedic Department, South Western Sydney Local Health District, Liverpool, NSW, Australia
[d] Whitlam Orthopaedic Research Centre, Ingham Institute for Applied Medical Research, 1 Campbell St, Liverpool, NSW, Australia
[e] Whitlam Orthopaedic Research Centre
[f] SWS Clinical School, University of NSW

ARTICLE INFO

Article history:
Received 16 June 2013
Accepted 30 July 2013

Keywords:
arthroplasty
Oxford score
administration
telephone
postal

ABSTRACT

Telephone and postal methods of administration of the Oxford Knee Score (OKS) and the Oxford Hip Score (OHS) were compared on 85 and 61 patients undergoing total knee arthroplasty (TKA) and total hip arthroplasty (THA), respectively. The test for equivalence was significant for both the knee ($P < 0.001$) and hip participants ($P < 0.001$) indicating that the modes of administration yielded similar results. The ICCs of the OKS and OHS were 0.79 (95% Confidence Interval (CI) 0.70, 0.86) and 0.87 (0.79, 0.92) respectively. The 95% limits of agreement were wide for both scores (OKS LOA, −8.6, 8.2; OHS LOA, −7.7, 5.3). The two modes of administration of the OKS and OHS produce equivalent survey responses at a group level but the same method of administration should be constant for individual monitoring in a clinical setting.

The Oxford Knee Score (OKS) and the Oxford Hip Score (OHS) are widely used patient-reported joint specific surveys, originally developed to evaluate success after total knee arthroplasty (TKA) and total hip arthroplasty (THA) [1]. Whilst the OKS and OHS have been shown to have acceptable face and construct validity, and have shown promising results in reliability and sensitivity to change [2–4], whether results vary depending on mode of administration – specifically, telephone interview or postal survey – remains unknown. The postal approach is relatively cheap, does not require simultaneous availability of both researcher and respondent, and circumvents the need for lengthy telephone interactions. For follow-up of large cohorts and in the absence of resources, the postal approach presents a viable option. However, postal surveys lack the immediate presence of the interviewer, which in itself, can assist survey completion as the interviewer can clarify ambiguous questions, probe for answers and ensure completeness of the survey.

Presently, there are minimal data concerning the effects of administration methods on the responses to the Oxford scores. It has been suggested by Whitehouse et al [5] that the Oxford scores are not ideal as a postal questionnaire due to the ambiguity of some of the questions. This was refuted by the original developers of the Oxford score as they maintained that in the original study only a minority required assistance in completing the questionnaire and that most of the surveys were in fact administered via post [6].

The literature presents mixed results concerning the impact of telephone versus postal administration on patient-reported outcome surveys. Dillman et al [7] found that oral modes of administration, such as telephone, are more likely to result in more extreme positive results compared to visual modes, such as mail, due to the factor of 'wanting to please'. Wood and McLauchlan [8] found differences between administration methods in the more general SF-36 but not the OHS, although their study did not have the same patients completing both methods of administration to allow direct comparison. A systematic review concerning questionnaire administration concluded that even though the literature was inconsistent, different modes of administration are likely to affect the quality of the data collected, with the biggest differences being between interview and self-administration methods [9].

The Oxford Scores have proven to be reliable and practical tools for patient-reported evaluation of arthroplasty surgery. The increasing volume of surgery [10] requires efficient means of data collection if patient-reported outcomes are to be a routine part of evaluation. Consequently, evidence concerning whether postal survey is an efficient method of survey administration is required to inform how best the data can be collected.

The primary aim of this study was to determine whether the two modes of administration – telephone or postal – were equivalent on a

The Conflict of Interest statement associated with this article can be found at http://dx.doi.org/10.1016/j.arth.2013.07.047.

Reprint requests: Marena Abdel Messih, MBBS Student, Liverpool Hospital, Orthopaedic Department, Locked Bag 7103, Liverpool NSW 1871.

0883-5403/2903-0009$36.00/0 – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.arth.2013.07.047

cohort level. A secondary aim was to determine if the surveys could be interchanged on an individual level for routine monitoring in daily clinical practice.

Materials and Methods

Patients undergoing unilateral knee or hip arthroplasty at a high-volume (> 500 procedures annually) arthroplasty centre were screened and approached during their preoperative education session or at the preadmission clinic between August 2011 and February 2012.

Inclusion criteria included patients who were receiving unilateral knee or hip arthroplasty, were 18 years or over, and were able to comprehend the protocol. Exclusion criteria included those unavailable for follow-up or who could not read or speak English. Patients willing to participate were required to provide written consent.

After enrolment into the study, each participant was randomly allocated the order of survey completion (telephone first, mail second ['Phone First'], or vice versa ['Mail First']) through a computer-generated sequence. A single interviewer was used and a telephone script followed to maintain consistency in what was communicated between participants. Each method of administration was completed within one week of the other to minimise time as a source of error.

Participants in the 'Mail First' group completed their survey unassisted on the day of enrolment (on site) and a telephone survey was undertaken one week later. Participants in the 'Phone First' group underwent a telephone survey within one day of enrolment. The same survey was then mailed out four days after the telephone survey. Reminder phone calls were made to participants who had not returned their mailed surveys within one week of the initial mail-out.

The Oxford scores consist of 12 disease-specific questions measuring patient-reported pain and function of the joint in question (hip or knee). Responses to each question are formatted as a 5-point Likert scale, scored from 0 (the worst outcome/most symptoms) to 4 (the best outcome / least symptoms). The scores for each question are added together for a final result between 0 and 48. Questions which required clarification (telephone) or were missed (postal) were recorded for audit purposes. Unanswered questions of the Oxford score were dealt with according to Murray et al [1], where an average was given to missing items but if more than 2 questions were left unanswered then the survey was excluded from the data. Any extra responses from the mail administration were also noted as well as which questions were missed to allow for a more coherent analysis of the differences between postal and telephone survey administration and its effects on the accuracy of the patients' results. If a question had 2 responses (i.e. the patient ticked 2 boxes) the most severe one was recorded.

Statistical Analysis

Using a standard deviation of 7 for both the OKS and OHS [1] and equivalence limits of $\pm$ 3 (assuming a clinically important difference in either Oxford score is greater than 3) [1], 2-sided alpha of 0.05, power of 0.8, and a conservative correlation of 0.6 between the two methods, a sample size of approximately 48 participants in each (joint) group was required. Assuming a 10% loss to follow up, a minimum of 54 in each joint group was required.

All analyses were conducted separately for hip and knee patients. Paired t-tests for equivalence were conducted to determine whether the Oxford scores between the two administration modes were similar using equivalence limits of -3 to 3. In equivalence testing, if the confidence interval for differences lies within the set boundary (here -3 to 3), this would be interpreted as equivalent. Internal consistency (reliability) of Oxford scores between modes of administration was analysed using the intraclass correlation coefficient (ICC). Co-efficients greater than 0.7 are recommended as adequate for group-based analyses [11] and 0.9 for individuals [12]. Repeated measures ANOVA was used to determine whether the order of administration resulted in significantly different Oxford scores whilst adjusting for patient characteristics (gender and any significant comorbidities) known to influence the scores [13,14]. The 'sequence of administration' variable was forced into the models regardless of P value, however, gender and comorbidities were only considered if they were associated with Oxford scores with $P < 0.2$ in univariate analysis. These variables were subsequently removed if in multivariate analysis their P values were greater than 0.05. In all definitive analyses, we considered $P < 0.05$ to be statistically significant. Bland and Altman 95% limits of agreement (B-A 95% LOA) [15] were also derived to establish whether the two modes could be interchanged within the same individual. In other words, the distribution-based analyses described above would determine whether the different modes are interchangeable within a group, but the B-A 95% LOA would describe whether the modes produced sufficiently reproducible scores such they could be interchanged within an individual when monitoring Oxford scores in an individual across time. The 95% LOA are calculated by: mean paired difference between modes $\pm$ 1.96 $\times$ (standard deviation of the mean paired difference). The analysis was generated using SAS Enterprise Guide software, Version 5.1 of the SAS System for Windows, Cary, NC, USA.

Results

Cohort ascertainment and retention are summarised in Fig. 1. Two hundred and fifty-one people booked for a TKA or THA at the Centre were screened over a six-month period. Many were excluded (n = 94), primarily due to language (n = 77) and some were lost to follow-up (n = 11), leaving 146 participants (n = 85 knees, n = 61 hips) for analysis. The rate of surgery was considerably slower for hip patients, hence a gross imbalance in the numbers of hip and knee participants recruited was evident over the six months.

All participants completed both survey modes prior to their surgery and only a minority required reminder phone calls to return their postal survey 1 week after their phone survey, with

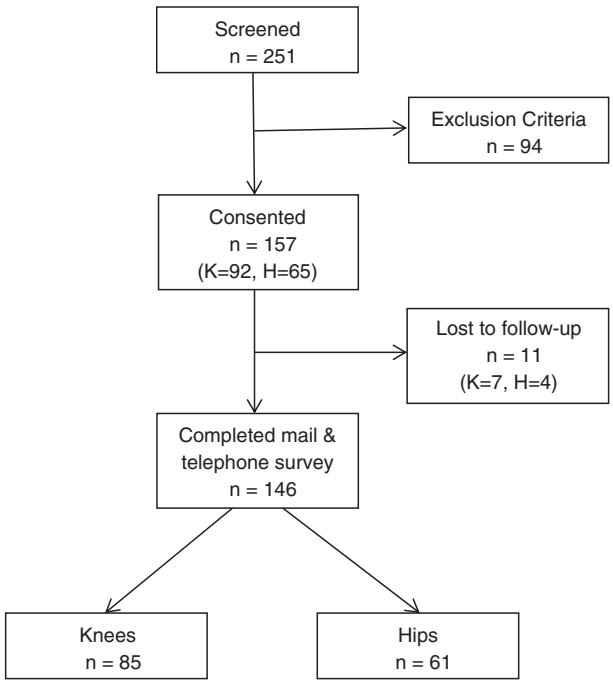

Fig. 1. Flowchart outlining patient recruitment and retention.