

Using Patient-Reported Outcome Measures (PROMs) Routinely: An Example in the Context of Elective Shoulder Surgery

Jill Dawson^{*1,2}, Katherine Rogers¹, Helen Doll¹, Ray Fitzpatrick¹, Cushla Cooper³ and Andrew Carr³

¹Department of Public Health, University of Oxford, Old Road Campus, Oxford OX37LF, UK

²Oxford Brookes University, Oxford, UK

³Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Science, University of Oxford, Windmill Road, Oxford OX37LD, UK

Abstract: *Objectives:* This paper evaluates the usefulness of a set of PROMs data collected prospectively, in a relational database, over several years. The data were collected as part of routine practice, to audit patients undergoing shoulder surgery.

Methods: Data, rendered anonymous, which included all upper limb surgical procedures and all outcome questionnaires (including Oxford Shoulder Score, OSS) data, were downloaded, prepared and analysed to produce the shoulder surgery study population. Details of procedures and questionnaires were merged by shoulder and analysed for completeness and accuracy. Pre- and post-operative OSS scores, with repeated observations from individual patients, were analysed using maximum likelihood mixed effects linear regression models.

Results: Data preparation and cleaning was intensive. The final dataset contained 2110 questionnaires representing 815 surgical procedures and 755 patients. In relation to procedures, only 538 (66%) had a pre- and post-operative OSS score: 78 (9.5%) a pre-operative OSS only, 199 (24.4%) a post-operative OSS only, and 31 (3.8%) no OSS completed. OSS questionnaires had been completed in varying numbers and at varying times per procedure. There was a considerable amount of 'missing' data that was not missing completely at random. Missing data had a significant influence on OSS scores.

Conclusions: In the absence of a research question (eg. exploratory research, descriptive audits, registers), a reason for collecting PROMs data should be stipulated and methods of data collection and storage standardised. Poor data cannot be 'fixed' in statistical analysis; statistical advice should be sought during the planning stages.

Keywords: Patient-reported outcomes, research methods, longitudinal study, shoulder, routine data.

INTRODUCTION

The use of patient-reported outcome measures (PROMs) is increasing. Most PROMs are questionnaires designed to assess outcomes of health care interventions in clinical trials [1] with analysis focused on the amount of change that has occurred in the patients' condition following treatment. Within health care, however, the routine use of generic and condition specific PROMs has also become widespread at a local level [2]. This includes the use of these measures in the contexts of audit and 'registries', to inform individual care and manage the performance of health care providers [2-6]. In the UK, PROMs have also been introduced throughout the National Health Service (since April 2009), to measure and improve clinical quality [7] in a number of areas of health care (including hip and knee arthroplasty, and hernia repair), which may further encourage their use at a local level.

The primary purpose of this paper is intended to be educational, particularly for individuals wishing to use PROMs at a local level, where appropriate resources and

expertise may be limited. We recently published [8] advice of a general nature on this theme, while this paper serves to illustrate some particular pitfalls that users of PROMs may encounter, by evaluating the usefulness and scientific merit of a real set of PROMs data that was collected at an orthopaedic centre prospectively and routinely over several years, in patients undergoing shoulder surgery. In evaluating these data, particular attention is focused on the following issues: 1) the accuracy of the data held in the database and their suitability for statistical analysis; 2) the extent to which data were missing and the nature of any apparent non-response bias; and 3) the factors related to post-surgical scores and the potential for missing data to affect the collected PROMs data (Oxford shoulder scores (OSS)[9]) pre- and post-surgery.

Recommendations follow, with general applicability, for those wishing to collect and monitor patient-reported outcomes over time, particularly where future statistical analysis of these data is envisaged.

MATERIALS AND METHODOLOGY

Institutional approval was obtained (and continues) in support of the ongoing orthopaedic, upper limb service evaluation.

*Address correspondence to this author at the Department of Public Health, University of Oxford, Old Road Campus, Oxford OX3 7LF, UK; Tel: 01865 289423; Fax: 01865 289436; E-mail: Jill.Dawson@dphpc.ox.ac.uk

Data Collection - The Database

Beginning in 1993, a relational database (MS Access™ software) was designed to be used at a specialist orthopaedic centre to audit upper limb surgery.^a The database is dynamic, in that data are continually added; it gradually evolved to fulfil a number of functions. Initially these included: (i) monitoring the process and pathway of care and clinical caseload of the surgical team; (ii) generating letters and discharge summaries to aid communications between the surgical team, patients and primary care; and (iii) generating reports based on aggregated data (e.g. numbers of patients treated per surgeon, rates of events etc over particular time periods). Subsequently, the database was adapted in order to monitor outcomes from shoulder surgery using the patient-reported Oxford Shoulder Score (OSS [9]). Thus, the database was intended to contain detailed information (demographics, diagnoses, surgical procedures) routinely entered on all patients undergoing shoulder surgery at this centre, with OSS data collected pre-operatively and at around 6 months, 12 months and 3 years post-surgery. The collection and entering of pre-operative OSS questionnaires first began in January 2000, with post-operative OSS questionnaires first collected from October 2001. Clinical details were entered by surgeons, while OSS data were collected and entered by two research nurses employed at different times (and not continuously). Pre-operative questionnaires were given to patients attending out-patient appointments and pre-admission clinics, with post-operative questionnaires either given to patients when they attended an out-patient appointment or mailed to their home.

The Oxford Shoulder Score (OSS)

The OSS is a validated, patient-reported, shoulder-specific questionnaire designed to measure pain and function in the context of shoulder surgery [9]. It is appropriate for all shoulder procedures except those for instability (which is more appropriately assessed using a different condition specific measure [12,13]). The OSS contains 12 items, each with five response options, scored from 0 to 4, with 0 representing greatest severity.^b Scores from individual items are summed to produce a single composite scale from 0 to 48, where a score of 0 represents the greatest disability.

2006 Cross-Sectional Postal Survey

A brief cross-sectional postal survey was conducted of all patients who had undergone shoulder surgery, excluding patients with shoulder instability problems. Sent out between June and October 2006, the survey questionnaire stated which shoulder (right or left) was the focus of interest, then

^aRelational databases use multiple tables to store data, with a mechanism for relating the tables to each other. Within each table, data are organised into fields (columns) and records (rows). Using relational databases is somewhat more complex than using spreadsheets and they do not have the broad range of calculating functions present in spreadsheets. However, spreadsheets and standard databases are not well suited for statistical analysis, so have to be exported into other statistical applications that generally have a 'flat file' structure [10,11].

^bThe original method scored each item from 1 to 5, from least to most difficulty or severity. Items were summed to produce a single score with a range from 12 (least difficulties) to 60 (most difficulties). The recommended method of scoring was recently revised to a 0 to 48 score to make the system more intuitive. The method of scoring used should always be stated in any study employing the OSS to avoid confusion.

presented the 12 OSS items. The questionnaire was accompanied by a cover letter and pre-paid envelope. Where necessary, a reminder letter was sent out after 2 weeks.

Statistical Analysis

Data were downloaded from MS Access™ in an anonymous form (individual patient's data entered into different database 'tables' were all linked by a unique identifier, but without the presence of names) and analysed within the SAS statistical package [14]. Downloaded data included all upper limb surgical procedures and all outcome questionnaire (including OSS) data which had been entered into one MS Access™ database ("questionnaires") table. Details of procedures (entered into a "procedures" table) and questionnaires were merged, by shoulder, in SAS. Tables and graphs were also produced using SAS. The final data download was on 22/01/2007.

In order to allow time for patient rehabilitation, OSS data were excluded if the completion date was recorded as < 3 months after the operation date. OSS data obtained beyond 4 years post-surgery (69 questionnaires) were also excluded, as these were known to have only been sent to a small proportion of patients.

Data are presented as N (%) or mean (SD) as appropriate. Group comparisons were performed using chi-square tests for categorical data, and t-tests or analysis of variance were used for continuous data. Post-operative OSS scores, with repeated observations from individual patients, were analysed separately and together, adjusting for potential confounders, using linear mixed effects models taking the patients as random effects. The non-linear effect of time was taken into account by adding both linear and quadratic effects as additional explanatory variables. The mixed models were fitted using the SAS procedure PROC MIXED. Factors that had a linear relationship with post-operative OSS score were fitted in both continuous and categorical form (by creating either three, four, or five categories of approximately equal size to maximise statistical power), and offered in both forms to the multiple variable model. Factors with a non-linear relationship with the OSS score were fitted in categorical form only.

Statistical significance was set at $p < 0.05$ throughout, with 95% confidence intervals used to express the uncertainty in the estimates.

Initial Data Preparation, Matching and Cleaning

Substantial data checking and preparation occurred prior to the final data download and statistical analysis. This process, which required ingenuity as well as complex computer programming, was conducted by a statistical analyst (KR) highly skilled in both MS Access™ and SAS analytical software. The period of data preparation took several months.

RESULTS

AIM 1: Accuracy of the Data Held in the Data Base and its Suitability for Immediate Statistical Analysis

On initial inspection of the data, various technical challenges were revealed, including: (i) OSS questionnaire and surgical procedure data, although linked by the patient's

representing a different denominator. Each perspective, however, differs from a patient-level analysis as an individual could undergo more than one procedure as well as complete more than one questionnaire (Fig. 1).

Sample Characteristics

The study sample, constituting data on 815 surgical procedures undertaken on 755 patients (see the Total column of Table 1), contained more female (N=441, 54%) than male (N=374, 46%) patients [note: patients contributed for each procedure they had; at the patient level N=403 (53%) female, N=352 (47%) male] with the overall mean (SD) age at surgery being 56.0 (13.3), range 17 to 87 years (patient level: age at first procedure = 55.9 (13.4), range 17 to 87 years). The majority of procedures (741, 90.9%) were classed by the surgeon (AC) as being major or complex, and 643 (78.9%) were performed as arthroscopic procedures. The mean (SD) pre-operative OSS score for the 616 procedures with pre-operative data was 23.60 (9.2), range 0 to 48. For the 571 patients with pre-operative data, the mean (SD) OSS score at first procedure was 23.78 (9.1), range 1 to 45.

Aim 2: The Extent to which Data were Missing and the Nature of any Apparent Non-Response Bias

Fig. (2) shows the frequency of routine OSS questionnaires collected per quarter annum from January 2000 until the end of 2006 together with the frequency of post-operative OSS questionnaires obtained in the cross-sectional survey.

The pattern of response shows that post-operative OSS data collection began tentatively, with the amount of data gradually increasing until early 2005. Thereafter, the frequency of data collection was erratic, including a six month period when post-operative data collection ceased altogether (when one research nurse left, and was replaced 6 months later), until Spring 2006, when 'routine' data

collection recommenced, but this was then eclipsed by the postal cross-sectional survey to all relevant patients.

Table 1 provides details of sample characteristics (demographic, operation type and extent, pre-operative OSS score), from the perspective of the procedure. Data are shown according to whether or not OSS questionnaires had been completed both pre- and post-operatively (n=538), only pre-operatively (n=78), only post-operatively (n=199), for all procedures (n=815), and where no OSS was completed (n=31).

Patients who completed the OSS only pre-operatively, or not at all, were significantly younger than those who completed the OSS either both pre- and post-operatively or only post-operatively, with 45% and 29% respectively aged <45 years (compared with 19% and 15% respectively; $\chi^2=34.6$, $df=3$, $p<0.0001$). Patients who completed the OSS both pre-and post-operatively were significantly more likely to be having arthroscopy (86% vs 63% overall), with those completing the OSS only post-operatively, or not at all, more likely to be having open surgery (40% and 48%, respectively vs 15% in those completing the OSS pre- and post-operatively or pre-operatively only; $\chi^2=70.2$, $df=3$, $p<0.0001$). This suggested that missing OSS data were not missing completely at random, with response being significantly associated with age, and type of operation.

Fig. (3) shows the frequency of OSS questionnaire completion by quarter of year before and after surgery for the 2110 questionnaires in the final 'questionnaires' dataset. The figure shows that the greatest proportion of questionnaires (N=316, 15%) were completed immediately before surgery. Post-operative OSS questionnaires were generally collected continuously throughout the follow-up period rather than at key time points (e.g. at around 6, 12, or 24 months).

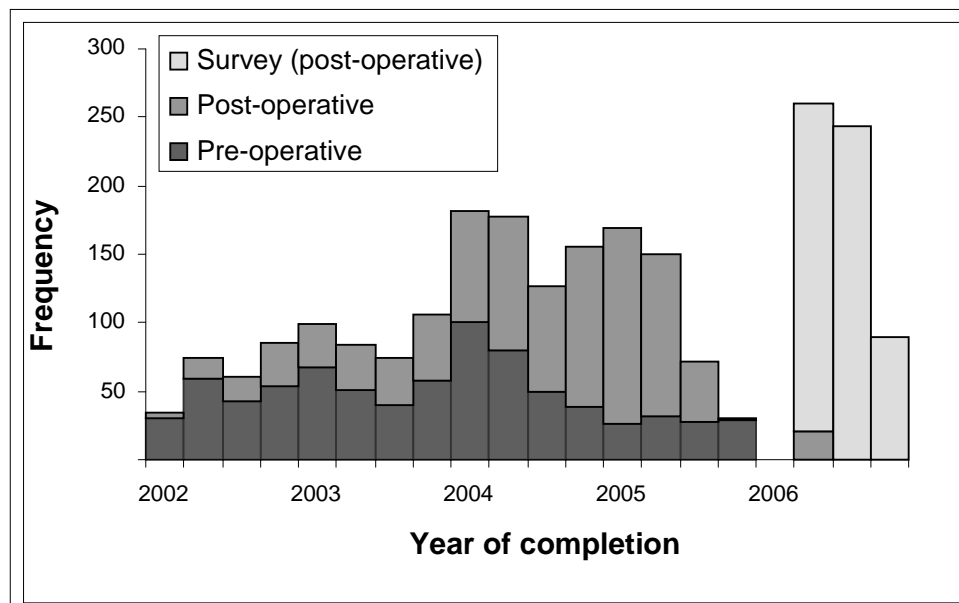


Fig. (2). Frequency within the total data set (N=2300 questionnaires, i.e. including the >1 extra 190 pre-operative questionnaires) of OSS questionnaires completed routinely per quarter (pre- and post-operative) from January 2002 to June 2006, plus the post-operative survey data during 2006.

Table 1. Sample Characteristics (Sex, Age, Operation Type and Extent, Pre-Operative OSS Score) Overall (for N=815 Procedures) and by OSS Completion Pattern: (i) Both Pre- and Post-Operatively (N=538), (ii) Only Pre-Operatively (N=78), (iii) Only Post-Operatively (N=199), and (iv) Not At All (N=31). Data are Shown as N (% within OSS Completion Group) Unless Otherwise Stated

Characteristic		(i) OSS Completed Pre- & Post-Op	(ii) OSS Completed Pre-Op Only	(iii) OSS Completed Post-Op Only	Total (i) + (ii) + (iii)	(iv) OSS Not Completed Pre- or Post-OP	
		N=538	N=78	N=199	N=815	N=31	
Sex	Female	292 (54.3)	36 (46.2)	113 (56.8)	441 (54.1)	14	45.2
	Male	246 (45.7)	42 (53.8)	86 (43.2)	374 (45.9)	17	54.8
Age (years)	<45	102 (19.0)	35 (44.9)	29 (14.6)	166 (20.4)	9	29
	45 to <60	246 (45.7)	26 (33.3)	69 (34.7)	341 (41.8)	10	45.5
	60 to <75	154 (28.6)	14 (17.9)	71 (35.7)	239 (29.3)	6	21.4
	75+	36 (6.7)	3 (3.9)	30 (15.1)	69 (8.5)	6	16.2
	Mean (SD)	55.6 (12.6)	48.5 (13.7)	60.1 (13.6)		54.5 (18.7)	
Operation type	Arthroscopic	464 (86.2)	59 (75.6)	120 (60.3)	643 (78.9)	16	51.6
	Other surgery	74 (13.8)	19 (24.4)	79 (39.7)	172 (21.1)	15	48.4
Extent of surgery	Major surgery	485 (90.1)	70 (89.7)	186 (93.5)	741 (90.9)	24	77.4
	not major	53 (9.9)	8 (10.3)	13 (6.5)	74 (9.1)	7	22.6
Pre-op OSS	Mean (SD)	23.6 (9.1)	23.8 (10.0)	--	--	--	

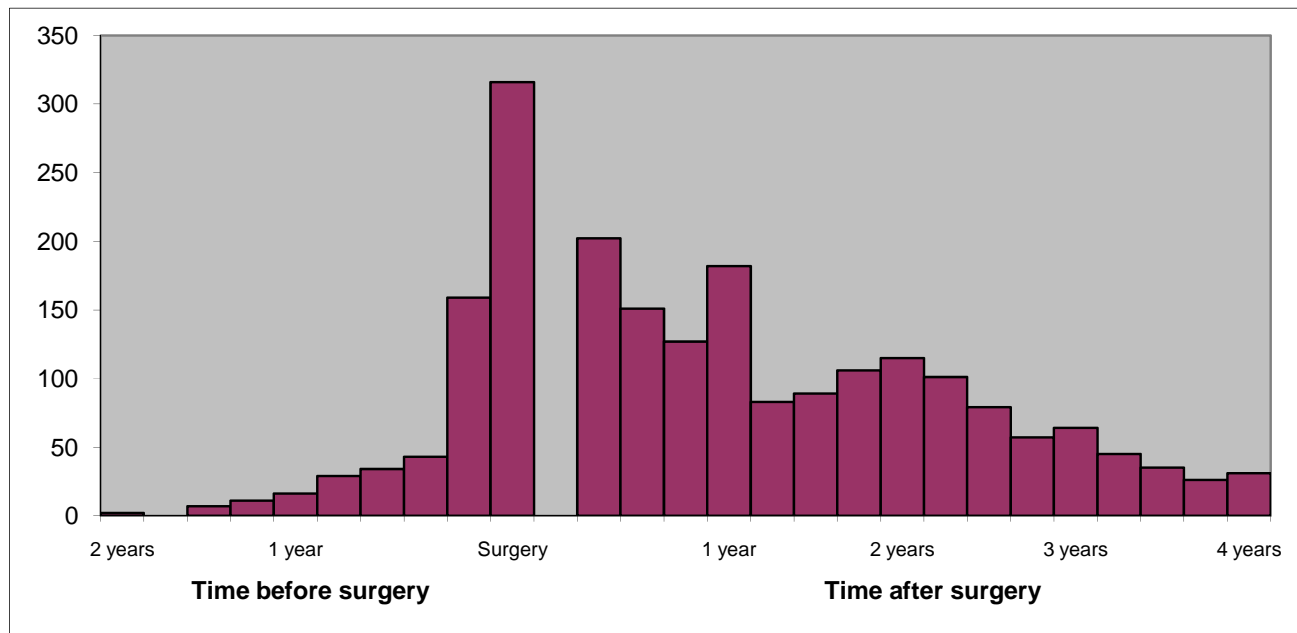


Fig. (3). Frequency of OSS questionnaire completion (for N=2110 questionnaires) in relation to time of surgery.

Table 2 presents details of the baseline characteristics of patients, comparing respondents versus non-respondents during several periods of follow-up. It is important to note that because follow-up was obtained continuously post-operation (see Fig. 3), the follow-up periods here were artificially constructed by dividing the follow-up questionnaires into 5 categories (follow-up time ±3 months) at the following times: 6 months (3 to <9), 12 months (9 to

<15), 18 months (15 to <21), 24 months (21 to <27), and 30 months (27 to <33). The majority of subjects did not provide OSS questionnaires during each period, with 37% providing data at 6 months, and 38%, 20%, 27%, and 19% at 12, 18, 24, and 30 months, respectively).

OSS data were under-represented for males and younger patients (Table 2). This is consistent with the data shown in Table 1. In addition, non-responders between 6 and 18

Table 2. Baseline Characteristics of Patients Undergoing Shoulder Surgery Since 1st January 2000 by Respondent (RESP) and Non-Respondent (NON-RESP) Status During Five Periods of Follow-Up Until 31st December 2006. Survey Data are Included. Data are Shown for the Final Dataset (N=815) and as N (% within Responder Category) Unless Otherwise Stated

Characteristic	Baseline	Follow Up Assessments									
	Pre-Op N=616	6 Months ±3 Months Resp N=225(36.5%) Non-Resp N=391(63.5%)		12 months ±3 months Resp N=234(38.0%) Non-Resp N=382(62.0%)		18 months ±3 Months Resp N=123(20.0%) Non-Resp N=493(80.0%)		24 months ±3 months Resp N=168(27.3%) Non-Resp N=448(72.7%)		30 Months ±3 Months Resp N=114(18.5%) Non-Resp N=502(81.5%)	
		Resp	Resp	Non-Resp	Resp	Non-Resp	Resp	Non-Resp	Resp	Non-Resp	Resp
Male	288 (46.7)	99 (44.0)	189 (48.3)	94 (40.1)	194 (50.7)	53 (43)	235 (47.6)	70 (41.6)	218 (48.6)	61 (53.5)	227 (45.2)
Age group (years):											
<45	137 (22.2)	33 (14.6)	104 (26.5)	31 (13.2)	106 (27.7)	18 (14.6)	119 (24.1)	33 (19.6)	104 (23.2)	19 (16.6)	118 (23.5)
45 to <60	272 (44.1)	105 (46.6)	167 (42.7)	114 (48.7)	158 (41.3)	61 (49.5)	211 (42.7)	78 (46.4)	194 (43.3)	50 (43.8)	222 (44.2)
60 to <75	168 (27.2)	70 (31.1)	98 (25.0)	67 (28.6)	101 (26.4)	37 (30.0)	131 (26.5)	48 (28.5)	120 (26.7)	37 (32.4)	131 (26.0)
75+	39 (6.3)	17 (7.5)	22 (5.6)	22 (9.4)	17 (4.4)	7 (5.6)	32 (6.4)	9 (5.3)	30 (6.6)	8 (7.0)	31 (6.1)
Age mean (SD)	54.7 (12.9)	57 (11.7)	53.3 (13.3)	57.5 (12.0)	53 (13.1)	55.8 (11.6)	54.4 (13.2)	55.3 (12.4)	54.4 (13.0)	57.1 (11.8)	54.1 (13.1)
Operation type:											
Arthroscopic	523 (84.9)	191 (84.8)	332 (84.9)	201 (85.8)	322 (84.2)	106 (86.1)	417 (84.5)	146 (86.9)	377 (84.1)	100 (87.7)	423 (84.2)
Operation extent:											
Major procedure	555 (90.0)	204 (90.6)	351 (89.7)	210 (89.7)	345 (90.3)	113 (91.8)	442 (89.6)	152 (90.4)	403 (89.9)	104 (91.2)	451 (89.8)
Pre-operative OSS score mean (SD)	23.6 (9.2)	23.1 (9.0)	23.9 (9.3)	22.4 (9.0)	24.3 (9.3)	22.4 (9.2)	23.9 (9.2)	24.1 (8.9)	23.4 (9.3)	24.7 (9.2)	23.3 (9.2)

months tended to have higher (i.e. better) pre-operative OSS scores (6 months: mean difference=0.80 (95% CI -0.71 to 2.31) $t=1.04$, $df=614$, $p=0.30$; 12 months: mean

difference=1.90 (95% CI 0.40 to 3.40) $t=2.49$, $df=614$, $p=0.013$; 18 months: mean difference=1.50 (95% CI -0.32, 3.32) $t=1.62$, $df=614$, $p=0.11$).

Table 3. Number of Post-Operative Questionnaires Completed by Respondent Characteristics (Sex, Age, Operation Type and Extent, Pre-Operative OSS Score). Data are Shown for All Procedures (N=815) and as N (% within Questionnaire Number Category) Unless Otherwise Stated

Characteristic	Number of Post-Operative OSS Questionnaires Completed by Patients				
	None (N=78)	One (N=278)	Two (N=254)	Three (N=131)	Four or More (N=74)
Sex					
Female	36 (46.2)	148 (53.2)	135 (53.2)	82 (62.6)	40 (54.1)
Male	42 (53.8)	130 (46.8)	119 (46.9)	49 (37.4)	34 (45.9)
Age at surgery (years)					
<45	34 (43.6)	70 (25.2)	43 (16.9)	12 (9.2)	7 (9.5)
45 to <60	27 (34.6)	118 (42.4)	98 (38.6)	72 (55.0)	26 (35.1)
60 to <75	14 (17.9)	63 (22.7)	88 (34.7)	40 (30.5)	34 (45.9)
75+	3 (3.9)	27 (9.71)	25 (9.8)	7 (5.3)	7 (9.5)
Mean (SD)	48.4 (13.6)	54.4 (14.1)	57.8 (13.1)	57.7 (10.3)	61.1 (11.2)
Operation type					
Arthroscopic	61 (78.2)	208 (74.8)	213 (83.9)	106 (80.9)	55 (74.3)
Other surgery	17 (21.8)	70 (25.2)	41 (16.1)	25 (19.1)	19 (25.7)
Extent of surgery					
Major surgery	69 (88.5)	247 (88.8)	232 (91.3)	119 (90.8)	74 (100)
not major	9 (11.5)	31 (11.2)	22 (8.7)	12 (9.2)	0 (0)
Pre-op OSS					
Mean (SD) (n)	23.8 (10.0) (n=78)	23.3 (8.9) (n=207)	24.3 (9.3) (n=175)	23.6 (9.8) (n=99)	23.1 (8.6) (n=56)

Further analysis focused on the number of post-operative OSS questionnaires completed (0, 1, 2, 3, or 4+), considered from the perspective of the procedure. Seventy eight (9.6%) procedures were associated with no post-operative OSS, 278 (34%) with 1, 254 (31%) with 2, 131 (16%) with 3 and 74 (9%) with 4 or more. Table 3 shows the distribution of sample characteristics (demographic characteristics, operation type and extent, pre-operative OSS score) by the number of post-operative OSSs.

The completion of a larger number of (or any) post-operative questionnaires was significantly associated with being older ($\chi^2=62.3$, $df=12$, $p<0.0001$), and non-significantly with having major surgery ($\chi^2=9.46$, $df=4$, $p=0.051$), having arthroscopic surgery ($\chi^2=7.82$, $df=4$, $p=0.099$), and with being female ($\chi^2=5.97$, $df=4$, $p=0.20$). The pre-operative OSS score was not associated with the number of post-operative questionnaires completed ($F=0.22$, $df=4,610$, $p=0.93$).

In summary, the dataset did not contain data collected at regular time intervals prior to and following surgery. Instead, OSS questionnaires were completed in varying numbers and at varying times. There was a considerable amount of 'missing' data with response being determined to some extent by patient characteristics such as age, gender, type and extent of surgery.

Aim 3: Factors Related to Post-Surgical OSS Score and the Potential for Missing Data to Affect Observed Patterns of OSS Scores Post Surgery

The OSS scores over time are shown in Fig. (4).

Scores generally decreased (i.e., become poorer) prior to surgery, increasing (i.e., improving) markedly following

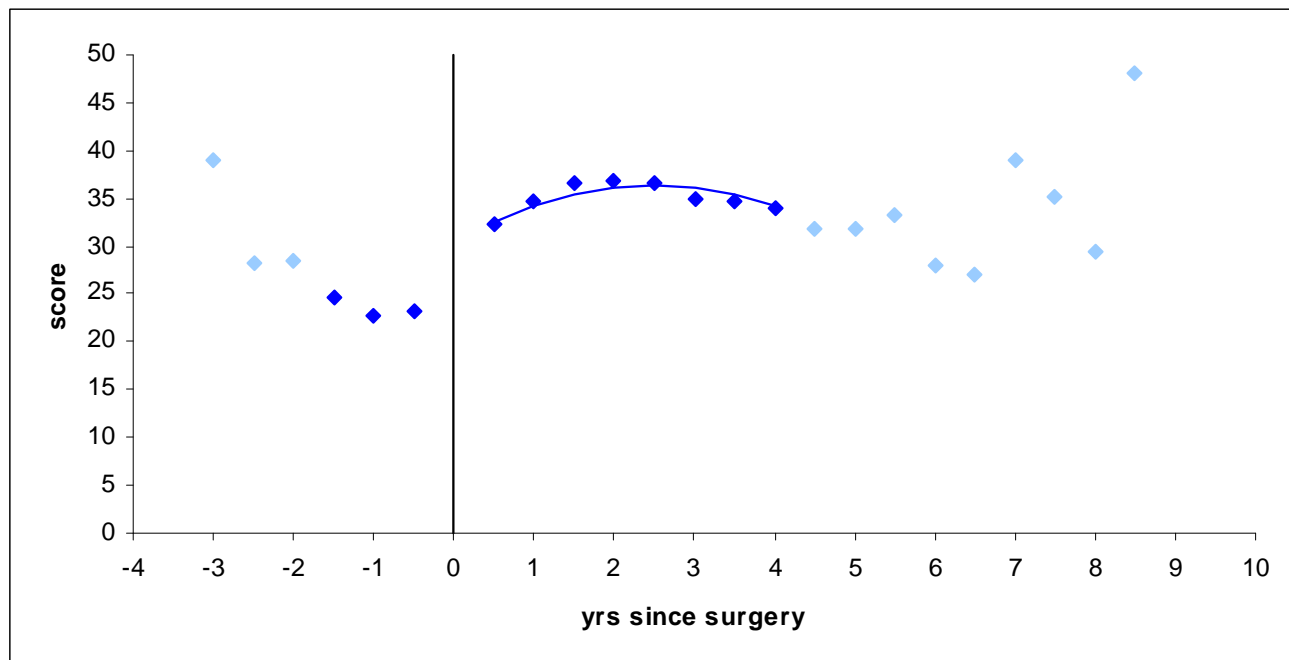
surgery and continuing to improve until around 2 years post surgery. However, these data had not been collected for each patient at each assessment time. It thus cannot be concluded that, for any one patient, such a pattern of OSS scores over time would be observed. The data shown in Tables 1-3 suggest (see Aim 2, above) that response (ie. presence of OSS data) was determined to some extent by patient characteristics such as age, gender, type and extent of surgery. Consequently, the observed pattern of OSS scores over time is likely to represent, at least partly, the effect of this differential response.

On modelling the relationship between each pre- and post-operative and surgical factor, and post-operative OSS score, many highly significant individual unadjusted associations were observed (all at least $p < 0.01$) (Table 4).

On multiple variable analysis (Table 4), the factors retaining independent effects with post-operative score were sex, age, pre-operative OSS score, timing of first post-operative OSS completion relative to operation date, first procedure, and extent of procedure. The number of post-operative OSS questionnaires completed ($F=2.45$, $df=1,226$, $p=0.12$) and the type of procedure ($F=2.44$, $df=1,226$, $p=0.12$) did not have a significant independent relationship with OSS score after adjusting for the other factors and were removed from the final model. The models were similar whether or not the effects of time and time squared were included; models unadjusted for time are presented.

DISCUSSION

This paper examined the usefulness and scientific merit of PROMs data that had been collected before and after shoulder surgery, for the purpose of routinely auditing



Line is time and time squared.

Pale dots represent data not used in the analysis (see statistical analysis section).

Fig. (4). Linear regression model (maximum likelihood mixed effects) of post-operative OSS scores, including repeated observations from individual patients, showing the trend of OSS scores over time (time and time squared).

Table 4. Individual Variable and Multiple Variable Models Obtained on Fitting Mixed Effects Linear Regression Models Showing the Relationships of Each Characteristic with Post-Operative OSS Score

Characteristic	Individual Variable			Multiple Variable		
	Overall F	β (SE)	p-Value	Overall F	β (SE)	p-Value
Sex	F _{1,229} =30.55		<0.0001	F _{1,227} =9.79		0.0020
Male (n=441)		1.00	-		1.00	
Female (n=374)		-4.72 (0.85)	<0.0001		-2.51 (0.80)	0.0020
Age group (4 categories)	F _{3,229} =13.96		<0.0001	F _{3,227} =5.56		0.0011
<45 (n=166)		1.00	-		1.00	
45 to <60 (n=341)		-0.76 (1.17)	0.52		-0.03 (1.07)	0.97
60 to 75 (n=239)		-2.40 (1.26)	0.059		-0.46 (1.15)	0.69
75+ (n=69)		-11.49 (1.91)	<0.0001		-7.11 (1.90)	0.0002
Pre-op OSS (continuous)	F _{1,228} =98.70		<0.0001			
		0.44 (0.044)	<0.0001			
Pre-op OSS group (3 categories)	F _{2,228} =40.48		<0.0001	F _{2,227} =25.67		<0.0001
<20 (n=202)		1.00	-		1.00	
20 to <30 (n=238)		5.20 (0.94)	<0.0001		4.50 (0.94)	<0.0001
30+ (n=175)		9.28 (1.04)	<0.0001		7.59 (1.07)	<0.0001
No. post-op OSS (continuous)	F _{1,228} =15.57		0.0001			
		-1.26 (0.32)	0.0001			
Time of first OSS (5 categories)	F _{4,228} =9.14		<0.0001	F _{4,227} =8.73		<0.0001
6 months (3 to 9 months) (n=298)		1.00	-		1.00	
1 year (9 to 15 months) (n=177)		3.62 (1.02)	0.0005		3.67 (0.91)	<0.0001
1.5 yrs (15 to 21 months) (n=113)		5.97 (1.27)	<0.0001		5.35 (1.13)	<0.0001
2 yrs (21 to 27 months) (n=81)		6.23 (1.49)	<0.0001		4.66 (1.35)	0.0006
≥2.5 yrs (27 to 33 months) (n=70)		4.04 (2.24)	0.073		2.23 (2.01)	0.2683
First procedure	F _{1,229} =19.03		<0.0001	F _{1,227} =20.17		<0.0001
Second/revision (n=55)		1.00	-		1.00	
First (n=760)		8.49 (1.95)	<0.0001		7.68 (1.71)	<0.0001
Type of procedure	F _{1,229} =17.22		<0.0001			
Other (n=172)		1.00	-			
Arthroscopy (n=643)		5.19 (1.25)	<0.0001			
Extent of procedure (3 categories)	F _{2,229} =12.12		<0.0001	F _{2,227} =5.09		0.0069
Minor (n=58)		1.00	-		1.00	
Inter/Major (n=628)		-2.43 (1.57)	<0.0001		-4.55 (1.43)	0.0017
Complex (n=129)		-9.11 (2.05)	0.12		-3.64 (1.96)	0.0647

clinical practice. The results are important because the routine use of PROMs is widespread in health care and interest in such routinely collected data continues to grow, particularly in the context of audit (and 'registers'), to inform individual care and manage the performance of health care providers.

PROMs have been used in clinical trials [15, 16], national audits [17], and registers for joint replacement [18] and other conditions [19]. PROMs are used to evaluate change in health status following health care interventions and offer many advantages over other methods of assessment

[1]. For results to be meaningful, however, in addition to using appropriate psychometrically validated measures, logistical and design features associated with data collection and storage also need to be specified, standardised and adhered to. This requires considerable care and planning. For instance, conditions and treatment that can involve bilateral structures (eg. joints, eyes, breasts), or that may require revision or repeated courses of treatment, can create complexity at every stage.

In clinical trials, the design of different study stages, including data collection time points, is determined by the

prior specification of a research question. The most effective trials involving PROMs are those that achieve very high questionnaire response rates at the pre-specified time points [20]. Nevertheless, procedures designed to maximise response rates can create extra costs [20, 21]. In contexts other than clinical trials, or other forms of funded longitudinal studies – such as audit or clinical governance – similar rules apply to data collection methods and storage if PROMs data are ultimately to be usable and not prove misleading. Data should be collected at pre-specified time points so that particular questions (e.g., the success of a procedure at one year post operation) can be addressed. It is important to recognise that collecting data continuously, but irregularly, after surgery will limit the usefulness of the data: more data does not necessarily mean better data.

The main findings of the analysis were firstly, that the data were found not to be suitable for immediate statistical analysis: considerable work and expertise was needed to increase the usability of the data. Secondly, the scope for any analysis of change in patients' shoulder problems was limited by incomplete pre- and post-operative outcomes data. Follow-up OSS data had also been obtained in a continuous manner throughout the collection period, rather than at particular junctures relative to patients' operation dates. This led to biases in the OSS data obtained, with male and younger patients, and those with non-major surgery being less likely to have completed questionnaires than other patients. In addition, individual patients had been allowed to contribute variable numbers of outcome questionnaires, with more questionnaires being completed by female and older patients, and those with major surgery, who also had poorer

post-operative scores. This likely occurred because OSS data had frequently been collected in out-patient departments, with patients having a poorer outcome being more likely to be seen more frequently and for a longer period following surgery.

The main consequence of the informal pattern of data collection, which led to a large amount of 'missing' pre-and post-operative OSS data, was the difficulty associated with using the data in any overall analysis of change in patients' health status following shoulder surgery. This limitation was highlighted by the regression modelling of the data which showed the effect that this known missing data had on the post-operative outcome scores (OSS). OSS scores depended strongly and independently on sex, age, pre-operative OSS scores, the time of first post-operative questionnaire, whether the procedure was a first procedure or not, the type of procedure, and the extent of the procedure. Thus, any change over time in the proportions of patients with each of these characteristics would affect the observed OSS score, with it thus being difficult to disentangle the effect of time from the effect of patient characteristic.

Our findings are remarkably similar to those reported following evaluation of the American Academy of Orthopaedic Surgeons' Musculoskeletal Outcomes Data Evaluation and Management System (MODEMS) initiative (beginning in 1997)[21], which aimed to collect PROMs from 1000 orthopaedists on consecutive patients undergoing hip or knee replacement. Our findings confirm the importance of adhering to a focussed process of data collection, with data being collected at regular and well-defined intervals.

- Poor data cannot be 'fixed' in an analysis by a statistician. Seek advice (or collaborate) from those with relevant expertise from the beginning.
- In the absence of a research question (eg. exploratory research, descriptive audits), a reason for collecting PROMs data, preferably with reference to an event eg. an intervention (with a date), & any follow-up period, should be stipulated before commencing data collection. This will help guide & standardise methods of data collection as well as the design of any associated database.
- If using a relational database, a unique patient identifier must be used to ensure that cross-sectional and longitudinal records are accurately linked. Other methods of linking patient-level data, such as linking questionnaires to procedures, should also be considered.
- For conditions (& interventions) that can affect bilateral structures (eg. joints, eyes, breasts), the unit of analysis (ie. patient versus (right or left) joint, eye or breast) should be decided in advance, & any database designed accordingly. If a treatment can be given repeatedly for the same condition over time, this too may affect database design - & data collection.
- Dates are crucial to longitudinal outcomes analysis, but some are more salient than others. Outcome measures need to be obtained & recorded with the date of completion (not the date of data entry), & with reference (labelled with, &/or linked) to the date of an intervention or event of interest.
- Follow-up outcomes data should be obtained once (& entered once) only, from all participants at a particular point (or points) in time following the intervention or event of interest. This can be aided by setting up automatic prompts to prevent duplicate entries.
- Systematic methods of data collection should be piloted & reviewed at an early stage. Once practicable methods have been devised they should be adhered to. These methods may need regular checking.
- Data need to be stored in a database or spreadsheet in a manner (unambiguous & appropriately labelled) that allows for immediate statistical analysis without the need for detective work & complex data programming. The aim should be to minimise complexity.
- Before too many cases (no more than 20) have been entered, methods for downloading data & conducting some simple analyses should be piloted.
- Every effort should be made to obtain complete follow-up data: the intensity with which follow-up information is sought & obtained is known to greatly influence response rate and thus study results [22]. Thus every effort should be made to deal with this, & other forms of missing data, & the biases that might otherwise occur.

Fig. (5). Key considerations, at the planning stage, before routinely collecting and storing PROMs data intended for future analysis.

CONCLUSION

Many of the issues identified by this evaluation arose because of simple omissions at the planning stage, before data were collected, and represent pitfalls which are easy to fall into, but which are also largely avoidable. For those wishing routinely to collect PROMs data to be analysed in the future, we suggest a number of key considerations which need to be taken into account at the planning stage (see Fig. 5).

FUNDING

No external funding was received in relation to this analysis/paper.

CONFLICT OF INTEREST

None of the authors have any conflict of interest in relation to this paper.

APPENDIX 1

1) Lack of Linkage Between OSS Questionnaires and Surgical Procedures

Each set of downloaded OSS item responses had been associated with a particular 'date of questionnaire completion'. None, however, had been directly linked with a particular surgical procedure. This feature of the data occurred because, 'pre-operative' OSS data had been entered before surgery had taken place, when the date and nature of subsequent surgery might have been uncertain. Thus, pre- and post-surgical OSS scores had to be identified as such retrospectively, by aligning the 'date of questionnaire completion' with patients' operation dates, and the side (right or left) of surgery. This latter detail was also essential because, over time, patients could have surgery on both shoulders, as well as revised operations. Where this information remained missing after data cleaning and updating, associated OSS data were excluded from the analysis. Duplicate questionnaire entries for the same date were also excluded.

2) The Recorded Presence of an OSS Questionnaire in the Absence of Data

By default, the database recorded the presence of a completed OSS questionnaire whenever someone opened the OSS data entry facility, even when no data had actually been entered. This feature suggested that more questionnaires had been completed than was the case.

3) Missing Individual OSS Item Responses

The database contained missing individual OSS items. If more than two out of the 12 OSS items were missing for a patient on a particular date the overall score for that questionnaire was excluded from the dataset. Where only one or two items were missing, the mean value of responses to all other items was imputed for each missing item.

APPENDIX 2

1) Early Stages of Data Cleaning and Preparation (Summarised in Fig. 1)

The first column in Fig. (1) represents surgical procedures. The database originally contained data for 1467

upper limb surgical procedures. After data cleaning and limiting the analysis to shoulder cases alone (but excluding shoulder instability) 1000 shoulder procedures remained (representing 899 patients; a patient could have more than one procedure) conducted during the period June 1993 to mid January 2007. Out of these, only 846 surgical procedures (conducted from January 2000, when OSS questionnaires were first used, until December 31st 2005) were included (representing 774 patients), to allow sufficient time (data were downloaded on 22/01/2007) to elapse for OSS follow-up questionnaires to be obtained. However, where procedures were matched to questionnaires, requiring the presence of at least one pre- or post-operative questionnaire (note: only one pre-operative questionnaire – the one closest to the operation date – was used per procedure), this reduced to 824 procedures representing 761 patients. Excluding 9 procedures with post-operative questionnaires completed either before 0.25 years or beyond 4.25 years of the operation, the number of procedures (including the 2006 survey data) reduced to 815, representing 755 patients. This is the final data set. The majority of these 815 procedures (538, 66.0%) had matched pre- and post-operative questionnaires (1 pre-operative and at least 1 post-operative questionnaire), while 78 (9.6%) had a matched pre-operative questionnaire only and 199 (24.4%) at least one post-operative questionnaire but no pre-operative questionnaire. Thus 616 of these 815 procedures had matched pre-operative questionnaires (538 + 78). In terms of the individual 755 patients, 571 (75.6%) had a pre-operative questionnaire. In addition, for 698 (92.5%) of these 755 patients, only one procedure was performed, with 55 patients having two procedures, 1 three and 1 four (note: 815 procedures in total: 698+110+3+4).

The second column of Fig. (1) summarises the data preparation procedures from the perspective of the *questionnaires*. The database held various questionnaires, of which 3744 were OSSs completed between January 2000 and December 31st 2006. Of these 3744 questionnaires, 2570 could be matched with a procedure, including 52 that were matched to 2 procedures - representing both a post-operative assessment for one and a pre-operative assessment prior to a subsequent procedure (on the same shoulder), thus increasing to 2622. The exclusion of post-operative questionnaires which were collected either too soon after surgery (253 questionnaires were collected within 3 months), or too long after surgery (69 questionnaires were collected after 4 years and 3 months) left 2300 OSS questionnaires collected for 815 procedures and 755 patients. More than half of these 2300 questionnaires (1456, 63.3%) were post-operative only (i.e., not a pre-operative questionnaire for a subsequent procedure; in total 1494 were post-operative), with 768 (33.4%) being pre-operative only (in total 808 were pre-operative), and an additional 38 both pre- and post-operative (i.e., providing 76 questionnaires in total). The exclusion of 190 pre-operative questionnaires, where more than one such was available (the one closest to the operation date was retained), left 2110 questionnaires, with 616 being pre-operative.

REFERENCES

- [1] Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998; 2(14): i-74.

- [2] Appleby J, Devlin N. Measuring success in the NHS. Using Patient-Assessed Health Outcomes to Manage the Performance of Healthcare Providers. Commissioned and funded by Dr Foster Limited. London: The King's Fund 2004.
- [3] Wasson J, Keller J, Rubenstein L, Hays R, Nelson E, Johnson D. Benefits and obstacles of health status assessment in ambulatory settings: the clinician's point of view. *Med Care* 1992; 30(Suppl): MS42-MS49.
- [4] Greenhalgh J, Long AF, Flynn R. The use of patient reported outcome measures in routine clinical practice: lack of impact or lack of theory? *Soc Sci Med* 2005; 60(4): 833-43.
- [5] Haywood K, Marshall S, Fitzpatrick R. Patient participation in the consultation process: A structured review of intervention strategies. *Patient Educ Couns* 2006; 63: 12-23.
- [6] Timmins N. Assessing patient care - NHS goes to the PROMS. *BMJ* 2008; 336: 1464-5.
- [7] Department of Health. Our NHS Our future: NHS next stage review - interim report 04/10/2007 http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_079077. London: 2007.
- [8] Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr A. The routine use of patient reported outcome measures in healthcare settings. *BMJ* 2010; 2010(340): c186.
- [9] Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg [Br]* 1996; 78: 593-600.
- [10] Simpson A, Olson E. Mastering Microsoft Access for Windows 95. 3 ed. San Francisco: Sybex 1996.
- [11] Woeltje KF, Wurtz R. Informatics in Healthcare Epidemiology. In: Mayhall CG, Ed. Hospital Epidemiology and Infection Control. 3 ed. Philadelphia, Lippincott: Williams & Wilkins 2004; pp. 197-206.
- [12] Dawson J, Fitzpatrick R, Carr A. The assessment of shoulder instability: the development and validation of a questionnaire. *J Bone Joint Surg [Br]* 1999; 81: 420-6.
- [13] Moser JS, Barker KL, Doll HA, Carr AJ. Comparison of two patient-based outcome measures for shoulder instability after nonoperative treatment. *J Shoulder Elbow Surg* 2008; 17(6): 886-92.
- [14] SAS analytical software release 9.1 Cary, NC 27513-2414, USA. SAS Institute 2006.
- [15] Silverman LR, Demakos EP, Peterson BL, *et al.* Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the Cancer and Leukemia Group B. *J Clin Oncol* 2002; 20: 2429-40.
- [16] Grant AM, Wileman SM, Ramsay CR, *et al.* Minimal access surgery compared with medical management for chronic gastro-oesophageal reflux disease: UK collaborative randomised trial (online first). *BMJ* 2008; 337:doi:10.1136/bmj.a2664.
- [17] Williams O, Fitzpatrick R, Hajar S, *et al.* Mortality, morbidity, and 1-year outcomes of primary elective total hip arthroplasty. *J Arthroplasty* 2002; 17(2): 165-71.
- [18] Malchau H, Garellick G, Eisler T, Herberts P. Presidential guest address: the Swedish Hip Registry: Increasing the sensitivity by patient outcome data. *Clin Orthop Relat Res* 2005; 441: 19-29.
- [19] Zanolli G, Nilsson LT, Stromqvist B. Reliability of the prospective data collection protocol of the Swedish Spine Register: test-retest analysis of 119 patients. *Acta Orthop* 2006; 77(4): 662-9.
- [20] Ganz PA, Gotay CG. Use of patient-reported outcomes in phase III cancer treatment trials: lessons learned and future directions. *J Clin Oncol* 2007; 25(32): 5063-9.
- [21] Saleh KJ, bershadsky B, Cheng E, Kane R. Lessons learned from the hip and knee musculoskeletal outcomes data evaluation and management system. *Clin Orthop Relat Res* 2004; 429: 272-8.
- [22] Bracken M. Reporting observational studies. *Br J Obstet Gynaecol* 1989; 96(4): 383-8.

Received: February 2, 2010

Revised: May 26, 2010

Accepted: May 27, 2010

© Dawson *et al.*; licensee *Bentham Open*.This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.