CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	The effectiveness of providing peer benchmarked feedback to hip replacement surgeons based on patient-reported outcome measures—results from the PROFILE (Patient-Reported Outcomes: Feedback Interpretation and Learning Experiment) trial: a cluster randomised controlled study
Authors	Boyce, Maria B.;Browne, John P.
Publication date	2015-07-31
Original Citation	BOYCE, M. B. & BROWNE, J. P. 2015. The effectiveness of providing peer benchmarked feedback to hip replacement surgeons based on patient-reported outcome measures —results from the PROFILE (Patient-Reported Outcomes: Feedback Interpretation and Learning Experiment) trial: a cluster randomised controlled study. BMJ Open, 5:e008325 .http:// bmjopen.bmj.com/content/5/7/e008325.abstract
Type of publication	Article (peer-reviewed)
Link to publisher's version	http://bmjopen.bmj.com/content/5/7/e008325.full - 10.1136/ bmjopen-2015-008325
Rights	This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/ - http:// creativecommons.org/licenses/by-nc/4.0/
Download date	2026-11-02 08:21:36
Item downloaded from	https://hdl.handle.net/10468/2221



University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Downloaded from http://bmjopen.bmj.com/ on August 4, 2015 - Published by group.bmj.com

BMJ Open

The effectiveness of providing peer benchmarked feedback to hip replacement surgeons based on patient-reported outcome measures—results from the PROFILE (Patient-Reported Outcomes: Feedback Interpretation and Learning Experiment) trial: a cluster randomised controlled study

Maria B Boyce, John P Browne

To cite: Boyce MB, Browne JP. The effectiveness of providing peer benchmarked feedback to hip replacement surgeons based on patient-reported outcome measures—results from the PROFILE (Patient-Reported Outcomes: Feedback Interpretation and Learning Experiment) trial: a cluster randomised controlled study. *BMJ Open* 2015;**5**:e008325. doi:10.1136/bmjopen-2015-008325

▸ Prepublication history and additional material is available. To view please visit the journal (http://dx.doi.org/10.1136/bmjopen-2015-008325).

Received 28 March 2015
Accepted 15 May 2015



Department of Epidemiology and Public Health, University College Cork, Cork, Ireland

Correspondence to
Maria B Boyce;
m.boyce@ucc.ie

ABSTRACT

Objective: To test whether providing surgeons with peer benchmarked feedback about patient-reported outcomes is effective in improving patient outcomes.

Design: Cluster randomised controlled trial.

Setting: Secondary care—Ireland.

Participants: Surgeons were recruited through the Irish Institute of Trauma and Orthopaedic Surgery, and patients were recruited in hospitals prior to surgery. We randomly allocated 21 surgeons and 550 patients.

Intervention: Surgeons in the intervention group received peer benchmarked patient-reported outcome measures (PROMs) feedback and education.

Main outcome variable: Postoperative Oxford Hip Score (OHS).

Results: Primary outcome data were available for 11 intervention surgeons with responsibility for 230 patients and 10 control surgeons with responsibility for 228 patients. The mean postoperative OHS for the intervention group was 40.8 (95% CI 39.8 to 41.7) and for the control group was 41.9 (95% CI 41.1 to 42.7). The adjusted effect estimate was −1.1 (95% CI −2.4 to 0.2, p=0.09). Secondary outcomes were the Hip Osteoarthritis Outcome Score (HOOS), EQ-5D and the proportion of patients reporting a problem after surgery. The mean postoperative HOOS for the intervention group was 36.2 and for the control group was 37.1. The adjusted effect estimate was −1.1 (95% CI −2.4 to 0.3, p=0.1). The mean postoperative EQ-5D for the intervention group was 0.85 and for the control group was 0.87. The adjusted effect estimate was −0.02 (95% CI −0.05 to 0.008, p=0.2). 27% of intervention patients and 24% of control patients reported at least one complication after surgery (adjusted OR=1.2, 95% CI 0.6 to 2.3, p=0.6).

Conclusions: Outcomes for patients operated on by surgeons who had received peer benchmarked PROMs data were not statistically different from the outcomes of patients operated on by surgeons who did not receive feedback. PROMs information alone seems to be insufficient to identify opportunities for quality improvement.

Trial registration number: ISRCTN 69032522.

Strengths and limitations of this study

- This is the first randomised controlled trial to examine the impact of providing surgeons with peer benchmarked patient-reported outcome measures (PROMs) feedback.
- The study employed a complex multiphase, multicentre design and the data was analysed using multilevel modelling to account for a lack of independence between observations and surgeon-level effects.
- Surgeons in the intervention arm received statistically meaningful feedback on their performance, which involved considerable effort in collecting data before randomisation.
- Patient recruitment proved difficult in some hospitals and not all patients were invited to participate. This reflects the considerable practical challenges involved in collecting PROMs on a routine basis across different treatment sites.
- The research explores the influence of PROMs feedback on the outcomes of one elective procedure.

INTRODUCTION

Patient-reported outcome measures (PROMs) are questionnaires that assess patients' health,

Downloaded from http://bmjopen.bmj.com/ on August 4, 2015 - Published by group.bmj.com

health-related quality of life and other health-related constructs.[1] As a result of concerns about the narrow focus of traditional outcome measures, such as mortality and clinician-defined morbidity, many countries are interested in embedding PROMs within larger initiatives to compare the performance of healthcare providers.[2–10] The National Health Service (NHS) PROMs Programme in England is the most advanced example of this approach.[2] Introduced in 2009, it mandates the collection of PROMs for all patients undergoing hip replacement, knee replacement, hernia repair and varicose vein surgery.[11] The data can be used by patients and purchasers to select an NHS Trust for their surgical procedure, and by Trusts to stimulate quality improvements.[2 12 13] PROMs have been posited as superior to other performance management tools, such as waiting time targets, because these are better aligned with the ultimate interests and motivations of clinicians—that is, the health and quality of life of patients.[2]

A 2013 systematic review found that the use of PROMs as performance management tools was not well investigated.[1] Only one study had investigated the impact of peer benchmarking of PROMs data. This provided PROMs feedback to primary care practices and found no impact on the health of community-residing patients compared with patients covered by control practices who received no feedback.[14] A recent time series analysis evaluated the impact of the English NHS PROMs Programme over the period 2009–2012. The study, which may have been biased by time-varying confounders, such as changes in resources, workforce composition and technology in the NHS over the study period, found no consistent positive effect on patient outcomes. This led the study authors to conclude that the mechanisms by which PROMs feedback are provided should be improved. One way to do this would be to provide individualised, peer benchmarked feedback to consultant surgeons with educational materials to guide their interpretation of the data.[15] Individual-level peer benchmarking is thought to stimulate an intrinsic desire in healthcare professionals to succeed relative to their peers, and to stimulate audit and research activities that identify the mechanisms by which performance can be improved.[16] For example, professionals who are discovered to have poor performance might learn from the practices of those with the best performance.[17] This innovation is supported by a 2014 systematic review of the qualitative literature which suggested that PROMs feedback would be more useful if it was delivered in a clear format, supported by educational materials, and tailored to the activity of individuals.[18]

As there is a clear need at this point in the implementation cycle for a definitive effectiveness study, we conducted a randomised controlled trial to test whether providing individualised peer benchmarked PROMs feedback and educational support to orthopaedic surgeons improves outcomes for patients undergoing hip replacement surgery.

METHODS

As the intervention was designed to improve the outcomes of patients undergoing hip replacement surgery by enhancing the performance of their surgeons, a cluster randomised controlled trial with a 1:1 allocation ratio of surgeons to an intervention or control arm was used.[19] Eligible participants were consultant orthopaedic surgeons in the Republic of Ireland. Only high-volume surgeons were randomised so that sufficient data for peer benchmarking could be collected within the study timetable. 'High volume' was defined as having responsibility for at least 100 primary hip replacement procedures per year. Patients were eligible for inclusion if they were under the care of participating surgeons; over 18 years of age; and undergoing an elective, unilateral, primary hip replacement procedure. Patients were excluded if they were incapable of completing a written questionnaire due to cognitive impairment, poor sight, or literacy/language comprehension problems.[11]

Intervention

The feedback intervention was designed to replicate the methods used in the NHS PROMs Programme, with the exception that feedback was provided at the level of the individual surgeon rather than the whole organisation, and was supported by educational materials. The feedback report was designed using the results of a qualitative study which explored professionals' preferences for metrics used to compare performance.[20] Each surgeon was provided with feedback about their patients' responses to the Oxford Hip Score (OHS).[21] The OHS is the disease-specific PROM used by the NHS PROMs Programme to measure the performance of NHS Trusts which provide hip replacement surgery.[11] It consists of 12 items on symptoms and functional status, and is summated to an overall score of 0–48, where higher scores represent a better outcome.[21] When drawing statistical comparisons of surgeons' performance, case-mix adjustment of the OHS was undertaken to account for the patients' preoperative OHS, age, sex and self-rated general health status.[22] Surgeons were also provided with feedback on the proportion of their patients that reported an overall improvement in their hip problem and the proportion that reported having at least one of four postoperative complications. Statistical comparison of surgeons' performance on these latter metrics were unadjusted for case-mix, following the approach taken by the NHS PROMs Programme.[22] The report presented to individual surgeons clearly demonstrated how each surgeon performed in comparison to the other surgeons in the trial, but the identity of other surgeons remained anonymous. The report was delivered to surgeons in the intervention group in January 2013 by post and email (see online supplementary appendix S1). In addition, a 9 min educational video session was produced by an expert on the interpretation of PROMs data (JPB), and was made available to

Downloaded from http://bmjopen.bmj.com/ on August 4, 2015 - Published by group.bmj.com

surgeons in the intervention arm by an email link to a dedicated website. The educational session described each outcome measure and explained how to interpret the graphs included in the report, such as how to identify statistically significant and clinically important differences. Surgeons in the feedback group were given a minimum of 1 month to reflect on their performance, view the educational materials supporting the feedback, and adjust their care practices as they saw fit. Surgeons in the control arm did not receive a feedback report or education, but were treated the same as surgeons in the intervention arm in all other respects.

Outcome measures

The primary outcome measure used to evaluate the effectiveness of PROMs feedback was the difference between the mean postoperative OHS of patients in the intervention and control groups who were operated on in the period after feedback was delivered to the intervention group of surgeons.

Secondary PROMs included the Hip Osteoarthritis Outcome Score (HOOS),[23] the EQ-5D,[24] and the proportion of patients reporting an allergy or reaction to a drug, urinary problems, bleeding or wound problems after surgery. The HOOS summary score ranges from 0 to 44 (best health status) and the EQ-5D ranges from −0.59 to 1 (perfect health). To deal with missing items, the mean response was imputed if no more than five items were missing on the OHS and the HOOS, and the mode response was imputed if no more than two items were missing on the EQ-5D.[11]

Recruitment procedure and data collection

The Irish Institute for Trauma and Orthopaedic Surgery sent a letter of invitation on behalf of the study team to all 90 of their members. Thirty surgeons identified themselves as willing and eligible to participate. The president of the institute identified an additional seven eligible surgeons who did not initially respond to the invitation letter. Five of these consented to participate; thus, 35 of the 37 high-volume hip replacement surgeons operating at the time of study recruitment in the Republic of Ireland were included.

Patient recruitment occurred in two phases. The first phase was used to generate PROMs feedback to surgeons and targeted patients who received surgery between May 2011 and June 2012. The second phase was carried out after surgeons had been randomised to the intervention or control arms, and targeted patients who received surgery between February 2013 and December 2013.

Identical consent and data collection procedures were used in both recruitment phases. Nurses and registrars identified and invited eligible patients prior to their operation in a preoperative assessment clinic, if available, or alternatively when patients were admitted for surgery. MBB provided training to the data collectors at each site to standardise procedures. Patients were told that the aim of the study was to find out about how they felt before and after their operation, and to evaluate whether this information was useful to surgeons. Staff involved in patient recruitment were asked to prospectively record the number of patients at their centre who were mistakenly not considered for participation, deliberately excluded due to ineligibility, invited to participate in the study, or declined to participate.

If patients consented to participate, they were asked to fill out a questionnaire prior to their operation and were informed that they would be sent a follow-up questionnaire by post to their home address 6 months after the operation by researchers from University College Cork. A reminder letter was sent 4 weeks after the initial postoperative questionnaire had been posted if a reply was not received within this time frame. Preoperative questionnaires included demographic questions on the patient's age, sex and duration of symptoms, the OHS, the HOOS, the EQ-5D, and a general health status item. Postoperative questionnaires included the same questions as the preoperative questionnaire plus questions on overall outcome and postoperative complications.

Sample size

This trial required separate sample size calculations for the prerandomisation and postrandomisation phases of the study. The first calculation established the number of patients required to accurately benchmark surgeons for the feedback intervention. As estimates for a group-level minimal important difference on the OHS vary between three and five points, we chose the average of these estimates to inform our sample size calculations.[11 25] We calculated that complete outcome data on 25 patients per surgeon would be necessary to detect a minimally important difference of four points in the OHS between the average score for one surgeon and the average score for all surgeons, with 80% power at the 5% significance level and assuming a SD in the postoperative OHS equal to 8. We inflated this to 32 patients per surgeon to allow for attrition during postoperative follow-up. This was set as the minimum recruitment target for each surgeon during the prerandomisation phase of the trial. The second calculation established the sample size required to detect a significant difference in outcome between patients treated by surgeons who had received PROMs feedback and those who were under the care of surgeons in the control arm. To account for within surgeon clustering, an upper value of 0.034 for the intraclass correlation coefficient was estimated from postoperative OHS data collected during the prerandomisation phase. As 21 surgeons achieved their recruitment target during the prerandomisation phase, the number of clusters was fixed at 21. Therefore, to detect a difference of four points in the postoperative OHS between the feedback and control arms of the trial, by assuming a SD value of 8 with 95% power at the 5% significance level, data on the primary outcome for 156 patients was required in each arm. We inflated the recruitment target to 203 patients for each

Downloaded from http://bmjopen.bmj.com/ on August 4, 2015 - Published by group.bmj.com

study arm to allow for a loss of 30% of patients to follow-up.

Randomisation and masking

An independent statistician at the Clinical Research Facility in Cork randomised the surgeons. The statistician received a list of surgeons with concealed identities from the authors. Randomisation occurred at the same time for all surgeons after the collection of data for the feedback intervention had ended. Surgeons were stratified according to the public or private status of the hospitals within which they practiced and whether their performance, as measured by the postoperative OHS, during the prerandomisation phase of the trial was above or below average.[26] A strata block size of two was generated using the Rand Corporation random number table. A starting point for reading the table was selected at random using the Stat Trek programme.

It was not possible to blind clinicians to their allocation as receipt or non-receipt of the feedback intervention could not be disguised. After randomisation, patients were not informed of the trial arm to which surgeons had been allocated. Those recruiting patients were also not informed of surgeon allocation, but may have discovered this information through interaction with the surgeon concerned.

Statistical analysis

A linear mixed-effects regression model was used to estimate the difference in the primary outcome between the intervention and control arms. The model assumed a fixed effect for the influence of PROMs feedback and a random effect for the influence of surgeon-level characteristics on the postoperative OHS. In the main analysis, we used data from all patients who had returned postoperative data during the postrandomisation phase, and we adjusted the effect of PROMs feedback for the influence of patient-level characteristics (age, sex, preoperative score and general health status). Similar methods were used to evaluate the effect of PROMs feedback on the secondary outcomes. Separate linear mixed-effects regression models were used for the HOOS and EQ-5D, and a logistic mixed-effects regression model was used for the proportion of patients reporting problems after surgery.

To assess the impact of non-response to the postoperative questionnaire, the preoperative characteristics of patients who did not respond were compared across arms. Three sensitivity analyses were carried out: the first to assess the impact of imputing missing OHS items on the estimate of the effect of feedback, the second to examine the impact of including the hospital identity as a random effect into the mixed-effects model, and the third to assess the impact of imputing the last OHS observation carried forward for patients lost to follow-up.

For all tests, we used a value of 0.05 for the level of significance. The results report means and ORs with 95% CIs.

RESULTS

Eleven surgeons were randomised to the intervention arm and 10 to the control arm. All participating surgeons were male and had been consultants for 10 years on average. Nine surgeons worked only in a public hospital, four worked only in a private hospital, and nine worked in both public and private hospitals. Surgeon characteristics were similar across the study arms (table 1). Surgeons in the intervention arm received feedback about the performance of all 21 included surgeons. These reports covered the outcomes of 624 patients who had been recruited in the prerandomisation phase and had completed preoperative and postoperative questionnaires. The proportion of patients operated on by the included surgeons during this phase, who were formally considered for inclusion by local data collectors, was estimated to be 54%. Two per cent of the patients who were considered for inclusion were deemed ineligible by local data collectors and 7% of those deemed eligible refused to participate once invited. Eighty-two per cent of patients who consented to participate and completed the preoperative questionnaires went on to return the postoperative questionnaires at 6 months after surgery (figure 1).

The baseline (prerandomisation) performance levels of surgeons in the intervention and control arms were similar (table 1). In the prerandomisation phase, the mean adjusted change in OHS from before surgery to 6 months after surgery for the 270 patients of surgeons who were later allocated to the control arm was 21.1 (95% CI 20.0 to 22.3) compared with 21.5 (95% CI 20.1 to 22.4) for the equivalent group of 321 patients in the intervention arm. The mean postoperative OHS recorded for all surgeons during the prerandomisation phase (range 36.8–44.0) was at least four points below the OHS ceiling of 48 indicating that, in theory, all had the potential to achieve a clinically important improvement in performance. Patients excluded during the prerandomisation phase due to their surgeons not reaching sufficient recruitment levels, or to their not returning postoperative questionnaires (n=262) were similar to those included in the study, when comparisons were possible (see online supplementary appendix S2). The characteristics of patients of surgeons who were eventually allocated to the intervention and control arms were similar (see online supplementary appendix S3).

After the 21 included surgeons had been randomised, a new cohort of 550 patients were recruited to the trial. Two hundred eighty-four patients were under the care of surgeons in the intervention arm and 266 patients were under the care of surgeons in the control arm (figure 1). Fifty-one per cent of patients treated by surgeons in the intervention arm and 58% of patients treated by surgeons in the control arm over the postrandomisation recruitment period were considered for participation in the study. Of these, 2% of the patients in both arms were considered ineligible for participation by local data collectors, and 7% were invited to participate but refused to give consent. Patient baseline

Downloaded from http://bmjopen.bmj.com/ on August 4, 2015 - Published by group.bmj.com

Table 1 Baseline characteristics of surgeons and patients included in the postrandomisation study phase

Characteristics (level)	Control group	Intervention group
Surgeon	*N=10*	*N=11*
Male, n	10	11
Public, n*	4	5
Experience, mean (SD)†	9 (2.7)	10 (2.8)
Baseline performance in OHS, mean (SD)	21.1 (9.7)	21.2 (10.4)
Patients covered by feedback report, mean (SD)	27 (4.6)	29 (6.9)
Patient	N=266	N=284
Age (years), mean (SD)	66.3 (11.1)	64.5 (11.8)
Male, n (%)	141 (53)	146 (51)
Health status, n (%)		
Excellent	32 (12)	30 (11)
Very good	90 (34)	97 (35)
Good	105 (40)	123 (44)
Fair	29 (11)	17 (6)
Poor	4 (2)	11 (4)
Duration of symptoms (years), n (%)		
<1	43 (16)	52 (18)
1–5	183 (69)	184 (65)
6–10	25 (9)	30 (11)
>10	15 (6)	18 (6)
OHS preoperative, mean (SD)	19.9 (8.3)	19.1 (8.5)
HOOS preoperative, mean (SD)	17.7 (7.6)	17.1 (7.9)
EQ-5D preoperative, mean (SD)	0.4 (0.3)	0.4 (0.3)

*Public refers to the number of surgeons working in a public hospital only.
†Experience refers to the number of years since the surgeon became a consultant.
HOOS, Hip Osteoarthritis Outcome Score; OHS, Oxford Hip Score.

characteristics were similar across arms (table 1). A postoperative response rate of 83% was achieved for the intervention group and 86% for the control group. All surgeons in the intervention arm received the feedback intervention and educational session, and all surgeons randomised remained in the study and were included in the trial analysis. The mean period from the time the feedback intervention was provided to the time the last patient was recruited was 38 weeks for surgeons in the intervention arm (range 19–49) and 36 for surgeons in the control arm (range 17–49). Intersurgeon variation in outcome as measured by the intraclass correlation coefficient was 0.06 for the OHS, 0.06 for the HOOS, 0.05 for the EQ-5D, and 0.05 for the percentage of patients reporting a problem after surgery.

Primary outcome

Table 2 presents the effect of the intervention on outcomes. The unadjusted mean postoperative OHS for all patients was 41.3 (95% CI 40.7 to 42.0). The unadjusted mean postoperative OHS for the intervention group was 40.8 (95% CI 39.8 to 41.7), and for the control group was 41.9 (95% CI 41.1 to 42.7). The adjusted mean difference obtained from the linear mixed-effects model was −1.1 (95% CI −2.4 to 0.2, p=0.09).

Secondary outcomes

Table 2 also presents the effect of PROMs feedback on the secondary outcomes. The unadjusted mean postoperative HOOS for all patients was 36.6 (95% CI 36.0 to 37.3). The mean postoperative HOOS for the intervention group was 36.2 (95% CI 35.2 to 37.2), and for the control group was 37.1 (95% CI 36.3 to 37.9). The adjusted effect estimate obtained from the linear mixed-effects model was −1.1 (95% CI −2.4 to 0.3, p=0.1).

The unadjusted mean postoperative EQ-5D for all patients was 0.86 (95% CI 0.84 to 0.88). The mean postoperative EQ-5D for the intervention group was 0.85 (95% CI 0.82 to 0.87), and for the control group was 0.87 (95% CI 0.85 to 0.89). The adjusted effect estimate obtained from the linear mixed-effects model was −0.02 (95% CI −0.05 to 0.008, p=0.2).

The unadjusted percentage of all patients who reported a problem after surgery was 26% (95% CI 22 to 30). The percentage of patients who reported at least one complication after surgery in the intervention arm was 27% (95% CI 21 to 33), and in the control arm was 24% (95% CI 19 to 30). The adjusted effect estimate obtained from the logistic mixed-effects model was 1.2 (95% CI 0.6 to 2.3, p=0.6).

Sensitivity analyses

When comparing the postoperative OHS in the intervention and control arms, results from the sensitivity analyses were similar to those in the main analyses. This was the case when missing items were not imputed (effect estimate for OHS=−1.1, 95% CI −2.4 to 0.2, p=0.09), when imputing values for patients that were lost

Downloaded from http://bmjopen.bmj.com/ on August 4, 2015 - Published by group.bmj.com

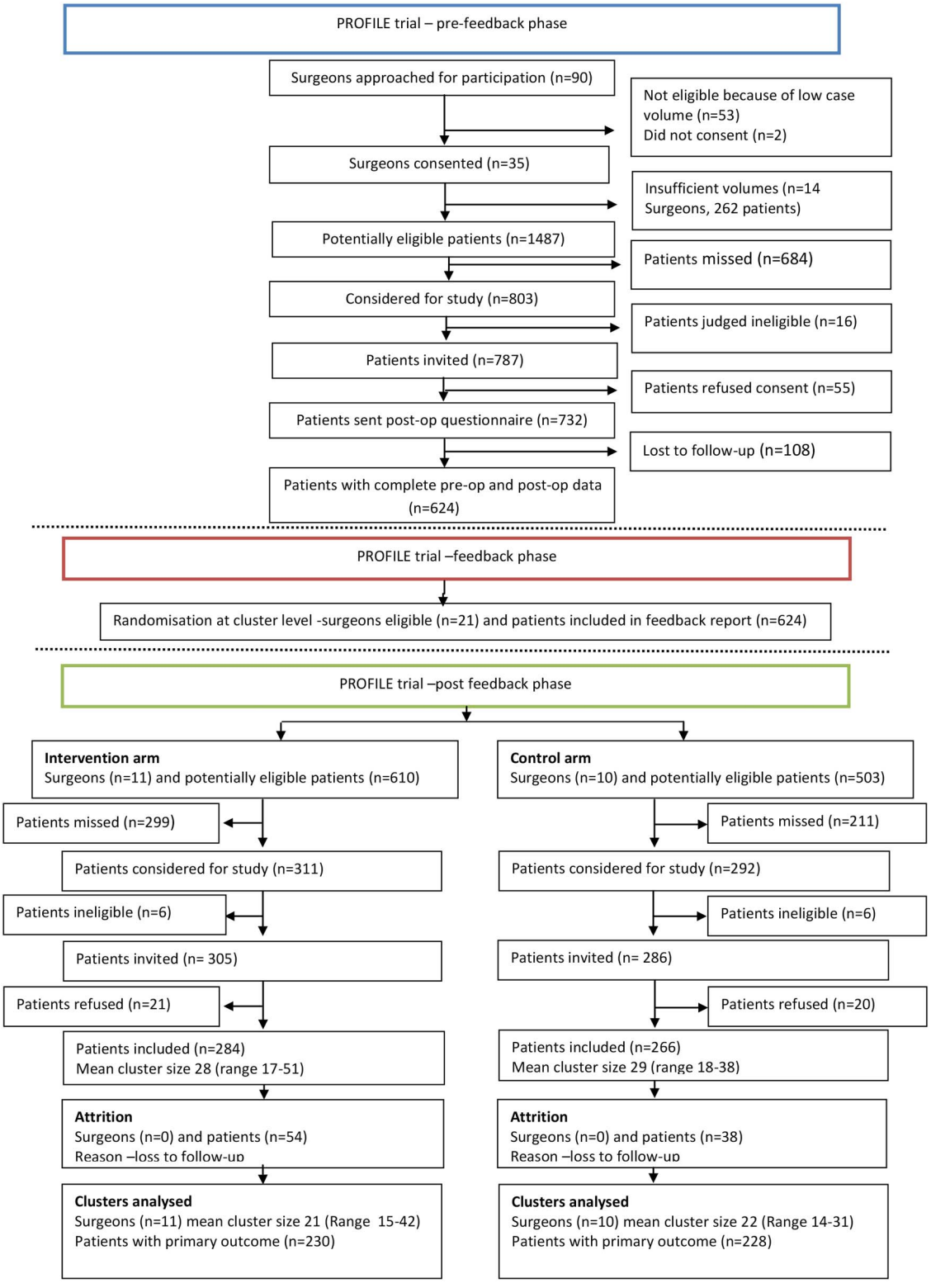


Figure 1 Flow of participants through the study (pre-op, preoperative; post-op, postoperative; PROFILE, Patient-Reported Outcomes: Feedback Interpretation and Learning Experiment).

to follow-up (effect estimate for OHS=−1.0, 95% CI −2.2 to 0.2, p=0.1), and when the hospital identifier was included as a random effect (effect estimate for OHS=−1.1, 95% CI −2.2 to 0.1, p=0.07).

DISCUSSION

This is the first randomised controlled trial to examine the impact of providing surgeons with peer benchmarked PROMs feedback. The study did not find a

Downloaded from http://bmjopen.bmj.com/ on August 4, 2015 - Published by group.bmj.com

Table 2 Primary and secondary outcomes

Outcome and period	Control group		Intervention group		Number of patients in multivariate analysis	Adjusted effect estimate* (intervention vs control) (95% CI)	p Value	ICC
	Number of patients	Mean (SD)	Number of patients	Mean (SD)				
Primary outcome								
Oxford Hip Score								
Baseline	266	19.9 (8.3)	284	19.0 (8.5)				
6 months	228	41.9 (6.4)	230	40.8 (7.4)	454	−1.1* (−2.4 to 0.2)	0.09	0.06
Secondary outcomes								
HOOS								
Baseline	261	17.7 (7.6)	280	17.1 (7.9)				
6 months	225	37.1 (6.4)	230	36.2 (7.7)	447	−1.1* (−2.4 to 0.3)	0.1	0.06
EQ-5D								
Baseline	253	0.43 (0.3)	270	0.38 (0.3)				
6 months	222	0.87 (0.2)	219	0.85 (0.2)	418	−0.02* (−0.05 to 0.008)	0.2	0.05
Proportion reporting problems after surgery (%)	229	24	230	27	455	1.2† (0.6 to 2.3)	0.6	0.05

*Estimates were obtained from a linear mixed-effects model adjusting for gender, age, health status, and baseline measure of outcome.
†Estimates were obtained from a logistic mixed-effects model adjusting for gender, age, health status, and baseline measure of outcome.
HOOS, Hip Osteoarthritis Outcome Score; ICC, intraclass correlation coefficient.

significant difference in outcomes for patients treated by surgeons who were randomised to a feedback group compared with patients treated by surgeons who were unaware of their performance.

Explanation of findings

A separately published qualitative study was undertaken to explore the views of the 11 surgeons in the intervention arm about the value of the feedback and educational support they had received.[18] The findings of this study help to explain the apparent ineffectiveness of the feedback intervention. Surgeons had differing opinions on the value of peer benchmarked PROMs data. Only two could be classified as strongly positive about the feedback they received and willing to take the information seriously when forming their model of care plans. Four surgeons were well disposed towards receiving patient feedback but were uncertain about the extent to which PROMs, such as the OHS or EQ-5D, could provide them with information that would outline specific improvements to patient care. The remaining five surgeons were sceptical about the value of PROMs feedback citing concerns about data validity, clinical relevance, and interpretability. Common to all surgeons were concerns about the high burden on local staff when recruiting patients and collecting data. This burden was considered to be a serious barrier to future engagement with PROMs data collection.[18]

Strengths and weaknesses of this study

A major strength of this study is the large amount of data that was collected to successfully deliver the feedback intervention. The study employed a complex multiphase, multicentre design, and expended considerable effort in collecting data before randomisation so that surgeons in the intervention arm received statistically meaningful feedback on their performance. We analysed the findings using multilevel modelling to account for a lack of independence between observations and surgeon-level effects. The possibility of performance and detection bias was unlikely as patients were blind to the allocation of the surgeon. All surgeons who were randomised remained in the trial and did not cross over; furthermore, patient response rates were high and similar across groups. We used a range of outcome measures which consistently found the same result, and we undertook qualitative interviews with surgeons in the intervention arm of the trial to gain a deeper understanding into why we did not find an intervention effect.[18] Finally, the study included 35 out of a possible 37 high-volume surgeons in the Republic of Ireland, and the model of care for hip replacement surgery in Ireland is similar to models used in other developed countries; thus, the external validity of findings is strong both within and outside Ireland. The study is only generalisable to high-volume surgeons, but there is evidence that there is no clinically important difference in the outcomes reported by patients operated on by high-volume and low-volume hip surgeons.[27]

Downloaded from http://bmjopen.bmj.com/ on August 4, 2015 - Published by group.bmj.com

The study also has some weaknesses. Patient recruitment proved difficult in some hospitals and not all patients were invited to participate. This introduces the chance of bias if the patients who were not recruited differed across the intervention and control arms of the study. This is an unknown risk for bias, but there is no obvious reason to suspect that it occurred. The recruitment levels observed in this study are similar to those observed in the NHS PROMs Programme[28] and reflect the considerable practical challenges involved in collecting PROMs on a routine basis across different treatment sites.[29 30] A further possible weakness is that the length of time between the receipt of feedback by surgeons in the intervention arm and the completion of post-feedback recruitment of patients subsequently treated by those surgeons may have been insufficient to capture the impact of potential improvements to care processes on patient outcomes. Furthermore, only one round of feedback was provided. Professionals may be more likely to engage with using PROMs data if they receive regular feedback reports and can observe consistent trends over time.[31] Finally, this research only explores the influence of PROMs feedback on the outcomes of one elective procedure and does not investigate the impact of extrinsic forces or motivations which can be employed alongside PROMs to improve care, such as the use of public reporting or pay for performance.[12]

Implications of findings for policy and practice

Our study adds to a growing body of evidence that embedding PROMs within quality assurance and improvement frameworks is unlikely to lead to patient benefit.[1 15] It is of direct relevance to those in charge of the English NHS PROMs Programme. The current practice of providing feedback at the NHS Trust level is highly unlikely to be effective if the intervention tested in this trial, which provides feedback to individual surgeons, is ineffective. This suggests that a more radical overhaul of current practice is required than simply changing the recipient of feedback. What can be carried out to improve the use of PROMs in this context? First, we need to question the assumption that improvements in the care of whole patient groups can be achieved by focusing on interprovider or interprofessional comparisons. For example, a recent study of hospital-level variation in PROMs for patients undergoing hip replacement, knee replacement, groin hernia repair or varicose vein surgery found 'little interprovider variation' which 'did not change significantly over time'.[15] In situations such as this, it may be more useful to focus on other aspects of the care episode, such as patient characteristics, type of treatment or type of provider, when trying to explain why some patients have better outcomes than others. Second, we need PROMs that are fit for purpose. If PROMs are to be used as diagnostic tools for poor clinical performance then evidence about their sensitivity and specificity in this context, validated against 'gold standard' measures of performance, is required. To date this evidence is not available for PROMs, such as the OHS, HOOS and EQ-5D, which is unsurprising given that these were not developed for this purpose. One obvious way to improve the usefulness of PROMs as quality improvement tools would be to allow patients the opportunity to record their own perceptions about care processes that either hindered or enhanced their journey to a full recovery. This information, if analysed alongside the outcome metrics generated by PROMs, would allow surgeons to identify where care might be improved. Third, audit and feedback interventions which use PROMs should be theoretically grounded and based on an explicit logic model.[32] PROMs feedback is unlikely to be useful when used in isolation and without an explicit implementation pathway to quality improvement. Performance monitoring can provide information about whether healthcare professionals perform better or worse than their peers but it does not explain why performance differs. In theory, the process of peer benchmarking assumes that professionals will be motivated to undertake additional audit or research activities to identify the reasons for differences in performance.[2 33] This assumes that professionals have the time, resources, knowledge, expertise, flexibility, and willingness to implement such activities. Capability to improve may be enhanced if professionals are provided with support to guide audit and research activities to identify areas for improvement.[15] For example, statistical and analytical support could be provided to link PROMs data to processes of care measures such as clinical data and patient experience data. Regular meetings between peers and researchers could also be facilitated where PROMs are discussed alongside other process and outcome data. These supports would allow for an in-depth examination of the care received by patients who report a poor outcome after surgery, and for peer learning about successful quality improvement initiatives. Opportunities to improve might also be enhanced if structural barriers to the implementation of quality improvement plans are removed.[34] This will require the flexibility to allow changes to patient pathways, to support additional investments in training, equipment and infrastructure, and to improve access to rehabilitation services if necessary.

CONCLUSION

The use of PROMs to peer benchmark surgeon performance on one occasion was not associated with superior patient outcome. Substantial changes to the way that PROMs are used in the quality improvement field may be needed if their full potential is to be realised.

Contributors MBB and JPB were involved in the conception and design of the study, data collection, and analysis and interpretation of data. Both authors were involved in drafting of the manuscript and revising it critically for important intellectual content, and approved the final version to be published. MBB is the guarantor.

Funding This work was funded by the Health Research Board in Ireland under Grant No. PHD/2007/16.

Downloaded from http://bmjopen.bmj.com/ on August 4, 2015 - Published by group.bmj.com

Competing interests None declared.

Ethics approval The Clinical Research Ethics Committee of the Cork Teaching Hospitals approved the study protocol, as well as the ethics committees in each participating hospital.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

REFERENCES

1. Boyce MB, Browne JP. Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients? A systematic review. *Qual Life Res* 2013;22:2265–78.
2. Devlin N, Appleby J. *Getting the most out of PROMs: putting health outcomes at the heart of NHS decision-making*. London: King's Fund, 2010.
3. Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013;346:f167.
4. Health Services Advisory Group. Medicare Health Outcomes Survey. USA, 2011 [17/11/11]. http://www.hosonline.org/
5. Nelson EC. *Using patient-reported information to improve health outcomes and health care value: case studies from Dartmouth, Karolinska and Group Health*. The Dartmouth Institute for Health Policy and Clinical Practice, 2012.
6. Callaly T, Hyland M, Coombs T, *et al*. Routine outcome measurement in public mental health: results of a clinician survey. *Aust Health Rev* 2006;30:164–73.
7. Meehan T, McCombes S, Hatzipetrou L, *et al*. Introduction of routine outcome measures: staff reactions and issues for consideration. *J Psychiatr Ment Health Nurs* 2006;13:581–7.
8. Pirkis J, Burgess P, Coombs T, *et al*. Routine measurement of outcomes in Australia's public sector mental health services. *Aust New Zealand Health Policy* 2005;2:8.
9. Canadian Institute of Public Health. *Health outcomes of care: an idea whose time has come*. Ottawa, Ontario: CIHI, 2012.
10. Delnoij DM, Westert GP. Assessing the validity of quality indicators: keep the context in mind! *Eur J Public Health* 2012;22:452–3.
11. Browne J, Jamieson L, Lewsey J, *et al*. Patient Reported Outcome Measures (PROMs) in Elective Surgery-Report to the Department of Health. London: London School of Hygiene and Tropical Medicine, Royal College of Surgeons of England, 2007.
12. Berwick DM, James B, Coye MJ. Connections between quality measurement and improvement. *Med Care* 2003;41(1 Suppl):I30–8.
13. Health and Social Care Information Centre. Patient reported outcome measures. UK, 2013. http://www.hscic.gov.uk/proms
14. Weingarten SR, Kim CS, Stone EG, *et al*. Can peer-comparison feedback improve patient functional status? *Am J Manag Care* 2000;6:35–9.
15. Varagunam M, Hutchings A, Neuburger J, *et al*. Impact on hospital performance of introducing routine patient reported outcome measures in surgery. *J Health Serv Res Policy* 2014;19:77–84.
16. Moriarty JP, Smallman C. En route to a theory of benchmarking. *Benchmarking* 2009;16:484–503.
17. Gawande A, ed. *Better: a surgeon's notes on performance*. London: Profile Books Ltd, 2007.
18. Boyce MB, Browne JP, Greenhalgh J. Surgeon's experiences of receiving peer benchmarked feedback using patient-reported outcome measures: a qualitative study. *Implement Sci* 2014;9:84.
19. MRC. *Cluster randomised trials: methodological and ethical considerations*. Medical Research Counil, 2002.
20. Hildon Z, Neuburger J, Allwood D, *et al*. Clinicians' and patients' views of metrics of change derived from patient reported outcome measures (PROMs) for comparing providers' performance of surgery. *BMC Health Serv Res* 2012;12:171.
21. Dawson J, Fitzpatrick R, Churchman D, *et al*. *User manual for the Oxford Hip Score (OHS). Information on development and use of the OHS*. Oxford: University of Oxford, 2010.
22. Department of Health. Patient Reported Outcome Measures (PROMs) in England. The case-mix adjustment methodology. England: Department of Health, 11 April 2012.
23. Davis AM, Perruccio AV, Canizares M, *et al*. Comparative, validity and responsiveness of the HOOS-PS and KOOS-PS to the WOMAC physical function subscale in total joint replacement for osteoarthritis. *Osteoarthritis Cartilage* 2009;17:843–7.
24. Cheung K, Oemar M, Oppe M, *et al*. *User Guide: basic information on how to use EQ5D*. EuroQol Group, 2009.
25. Beard D, Harris K, Dawson J, *et al*. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. *J Clin Epidemiol* 2015;68:73–9.
26. Altman DG. Randomisation. *BMJ* 1991;302:1481–2.
27. Varagunam M, Hutchings A, Black N. Relationship between patient-reported outcomes of elective surgery and hospital and consultant volume. *Med Care* 2015;53:310–16.
28. Lingard L, ed. *Power of the NHS patient reported outcome measures (PROMs) programme to improve quality*. Edinburgh: ISqua, 2013.
29. Jahagirdar D, Kroll T, Ritchie K, *et al*. Using patient reported outcome measures in health services: a qualitative study on including people with low literacy skills and learning disabilities. *BMC Health Serv Res* 2012;12:431.
30. McGrail K, Bryan S, Davis J. Let's all go to the PROM: the case for routine patient-reported outcome measurement in Canadian healthcare. *Healthc Pap* 2011;11:8–18; discussion 55–8.
31. Ivers N, Jamtvedt G, Flottorp S, *et al*. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;(6):CD000259.
32. Greenhalgh J, Pawson R, Wright J, *et al*. Functionality and feedback: a protocol for a realist synthesis of the collation, interpretation and utilisation of PROMs data to improve patient care. *BMJ Open* 2014;4:e005601.
33. Appleby J, Devlin N. *Measuring success in the NHS: using patient-assessed health outcomes to manage the performance of healthcare providers*. London: Dr Foster Ethics Committee and funded by Dr Foster Limited, 2004.
34. Eijkenaar F, Emmert M, Scheppach M, *et al*. Effects of pay for performance in health care: a systematic review of systematic reviews. *Health Policy* 2013;110:115–30.

Downloaded from http://bmjopen.bmj.com/ on August 4, 2015 - Published by group.bmj.com

BMJ Open

The effectiveness of providing peer benchmarked feedback to hip replacement surgeons based on patient-reported outcome measures —results from the PROFILE (Patient-Reported Outcomes: Feedback Interpretation and Learning Experiment) trial: a cluster randomised controlled study

Maria B Boyce and John P Browne

BMJ Open 2015 5:
doi: 10.1136/bmjopen-2015-008325

Updated information and services can be found at:
http://bmjopen.bmj.com/content/5/7/e008325

These include:

Supplementary Material Supplementary material can be found at:
http://bmjopen.bmj.com/content/suppl/2015/07/31/bmjopen-2015-008325.DC1.html

References This article cites 20 articles, 5 of which you can access for free at:
http://bmjopen.bmj.com/content/5/7/e008325#BIBL

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

Email alerting service Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections Articles on similar topics can be found in the following collections

Health policy (337)
Health services research (669)
Medical management (119)
Medical publishing and peer review (38)

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/

Downloaded from http://bmjopen.bmj.com/ on August 4, 2015 - Published by group.bmj.com

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/