

Title	The effectiveness of using patient-reported outcome measures as quality improvement tools
Authors	Boyce, Maria B.
Publication date	2014
Original Citation	Boyce, M. 2014. The effectiveness of using patient-reported outcome measures as quality improvement tools. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2014, Maria Boyce. - http://creativecommons.org/licenses/by-nc-nd/3.0/
Download date	2024-07-03 17:48:57
Item downloaded from	https://hdl.handle.net/10468/2038



UCC

Coláiste na hOllscoile Corcaigh, Éire
University College Cork, Ireland

The effectiveness of using patient-reported outcome measures as quality improvement tools

A thesis submitted to the National University of Ireland, Cork for the degree of Doctor of Philosophy in the Department of Epidemiology and Public Health, School of Medicine.

July 2014

Maria Boyce

BA, MA (Economics- health and evaluation studies)

Head of Department

Professor Ivan J. Perry

Supervisor

Professor John P. Browne

Table of Contents

List of Tables.....	iv
List of Figures	v
List of Appendices	vi
Glossary	vii
Declaration	viii
Dedication	ix
Acknowledgements	x
Thesis Abstract.....	xii
Chapter 1- Introduction & Background	16
1.1 Introduction	16
1.2 Background	16
1.2.1 Quality improvement in healthcare.....	18
1.2.2 Framework for quality improvement.....	19
1.2.3 Quality improvement strategies	20
1.2.4 Audit and feedback	23
1.2.5 Outcome measurement	27
1.2.6 Different uses of PROMs.....	28
1.2.7 The NHS PROMs Programme.....	30

1.2.8 Context of the research	43
1.3 Overview of the thesis	51
1.3.1 Research aim, hypothesis and objectives.....	51
1.3.2 Research methods	52
1.3.3 Clarification of the researcher’s role	53
1.3.4 Published papers to date	55
Chapter 2- Quantitative Systematic Review	57
2.1 Abstract	57
2.2 Introduction	58
2.3 Methods.....	66
2.4 Results	68
2.5 Discussion	83
Chapter 3- Qualitative Systematic Review	89
3.1 Abstract	89
3.2 Introduction	90
3.3 Methods.....	93
3.4 Results	95
3.5 Discussion	110
Chapter 4: The PROFILE Trial.....	116
4.1 Abstract	116
4.2 Introduction	118

4.3 Methods.....	119
4.4 Results	126
4.5 Discussion	134
Chapter 5- Qualitative Study.....	142
5.1 Abstract	142
5.2 Background	143
5.3 Methods.....	145
5.4 Results	152
5.5 Discussion	166
5.6 Conclusion	172
Chapter 6- Discussion	174
6.1 Summary of main findings.....	174
6.2 Implications of findings	183
6.2.1 Policy and practice	183
6.2.2 Research	188
6.2.3 Implications of the findings in Ireland	202
6.3 Limitations of the research.....	203
6.4 Conclusion	205
References	206
Appendices.....	225

List of Tables

Table 1: Categories of quality improvement interventions	21
Table 2: The journey of PROMs into policy in England.....	33
Table 3: Previous reviews on the effectiveness of PROMs feedback to healthcare professionals (n=7)	62
Table 4: Studies investigating the impact of PROMs feedback to healthcare professionals (n=17)	70
Table 5: Studies investigating the views of professionals (n=16)	97
Table 6: Taxonomy of themes, their definitions and excerpts from the studies	106
Table 7: Baseline characteristics between arms	131
Table 8: Primary and secondary outcome (baseline and 6 months)	132
Table 9: Characteristics of participants	152
Table 10: Themes, sub-themes and excerpts from the participants	154
Table 11: Mapping of themes and sub-themes across surgeons.....	162

List of Figures

Figure 1: Structure-process-outcome framework	20
Figure 2: Implementation pathway – audit and feedback of a process measure	25
Figure 3: Implementation pathway- audit and feedback of an outcome measure	26
Figure 4: Quality improvement pathways	41
Figure 5: Flow chart of studies in quantitative review	69
Figure 6: Flow chart of studies in the qualitative review	96
Figure 7: Flow of participants through the study	130
Figure 8: Example of peer benchmarked PROMs feedback	148
Figure 9: Conceptual framework of the relationship between themes	164
Figure 10: Overview of results, dilemmas and implications of findings.....	175

List of Appendices

Appendix 1: Quantitative Search Strategy	225
Appendix 2: Qualitative Search Strategy	229
Appendix 3: Critical appraisal of studies included in the qualitative systematic review using CASP236	
Appendix 4: Themes identified as barriers and facilitators to the use of PROMs within each study.	253
Appendix 5: Feedback report provided to surgeons in the trial.....	259
Appendix 6: Educational session provided to surgeons in the trial.....	260
Appendix 7: Training session provided to data collectors	261
Appendix 8: Information sheet provided to patients in the trial	262
Appendix 9: Study handbook provided to the data collectors.....	263
Appendix 10: Pre-operative questionnaire	264
Appendix 11: Post-operative questionnaire.....	265
Appendix 12: Pre-feedback differences in age, gender and scores between patients included in the study, excluded from the analysis and excluded from the study.	266
Appendix 13: Pre-feedback patient characteristics between arms.	268
Appendix 14: Discussion guide for interviews	270
Appendix 15: Examples of monthly newsletters provided to data collectors.....	274
Appendix 16: Published papers.....	275

Glossary

Term	Definition
BPT	Best Practice Tariff
BST	Basic Specialist Training
CIs	Confidence Intervals
COSMIN	Consensus-based Standards for the selection of health Measurement Instruments
CTT	Classical Test Theory
ED	Emergency Department
EQ-5D	EuroQoL-5 Dimension
GPs	General Practitioners
HIQA	Health Information and Quality Authority
HOOS	Hip Osteoarthritis and Outcome Score
HRQoL	Health-Related Quality of Life
HSE	Health Service Executive
IRT	Item Response Theory
LTT	Latent Trait Theory
NHS	National Health Service
OHS	Oxford Hip Score
PROFILE	Patient-Reported Outcome: Feedback Interpretation and Learning Experiment
PROM	Patient-Reported Outcome Measure
PICO	Patient, Intervention, Comparison, Outcome
QoL	Quality of Life
RCSI	Royal College of Surgeons Ireland
RCT	Randomised Controlled Trial
SF-36	Short Form-36
SHO	Senior House Officer
SMD	Standardised Mean Difference
UCC	University College Cork

Declaration

I, Maria Boyce, confirm that this thesis is my own work and has not been submitted for any other degree, either at University College Cork or elsewhere.

Signed: _____

Date: _____

Dedication

To Mam and Dad

Acknowledgements

To my supervisor Professor John Browne, thank you for introducing me to this area of research. I really appreciate your guidance throughout the past number of years. You always had an incredible ability to offer precise clarity and direction whenever I had a specific research query. In particular, I would like to express my gratitude to you for always striving for excellence. Thank you also to Dr. Joanne Greenhalgh who helped to enhance my qualitative skills, providing clear advice throughout the qualitative review and study. I would like to extend my appreciation to Dr. Carol Sinnott for providing me with feedback from a clinician's perspective during the analysis of my qualitative study. Thank you to the wider supports within Epidemiology and Public Health including Professor Ivan Perry, Professor Patricia Kearney, Dr. Birgit Greiner, Dr. Tony Fitzgerald as well as Vicky, Karen, Tara, Andrea and Dervla. I will always be grateful that I got a place on the HRB Scholars Programme. I would like to acknowledge all of those involved in setting up this Programme and to express sincere thanks for providing me with the opportunity to work among the experts in this field. I wish all my fellow scholars successful careers and collectively I hope we can build a greater capacity for Health Services Research in Ireland.

To Mam, thank you for your endless encouragement and support—your calming influence always helps to put problems into perspective. Thank you for teaching me to value what is really important in life. To Dad, you always promoted the importance of further education and you most definitely demonstrated the benefits of hard work—I know you would have been proud of this thesis. I owe a huge thanks to Suzanne who is the epitome of a caring big sister. Thank you for encouraging me through the more challenging times and for your artistic direction when designing

my questionnaires! I also wish to thank my family and friends for understanding my lack of commitment to life outside of work in the final months of my PhD. To my colleagues and good friends in Epidemiology and Public Health, you made the Department a truly positive environment to work within; without your support, good humour and outstanding work ethic, undertaking this PhD would surely have been a less fruitful and enjoyable experience. I will always remember the beautiful evenings we spent running and swimming in Kinsale to wipe away all the stresses of the day. A particular thanks to Sheena, Carol, Sarah-Jo, Martin, Margaret, Elsa, Rebecca, Jamie and Sharon for their help and support. Most importantly to my fiancé Michael, I would firstly like to apologise for all the lonely dinners you had to endure in the final phase of the study and secondly, I would like to sincerely thank you for all the hot cups of coffee, warm dinners, packed lunches, lifts to and from work and mind-clearing midnight walks. You not only made sure to keep my energy levels up, but also my spirits. Most of all, I wish to thank you for always finding a way to make me laugh at the end of a long day. It is my pleasure to finish this PhD so I can eventually begin planning our wedding!

Lastly, thank you to all those that participated in this research without whom this thesis would not have been possible.

Thesis Abstract

Aim: To investigate the value of using patient-reported outcome measures (PROMs) as a quality improvement tool.

Methods: Two systematic reviews were undertaken. The first review examined quantitative studies on the impact on patient outcomes of feeding back PROMs information to providers. The second review explored qualitative evidence on the barriers and facilitators to the use of PROMs in practice. These reviews informed the focus of the primary research. A mixed methods design was used to examine the impact of providing peer benchmarked PROMs feedback to consultant orthopaedic surgeons to improve patient outcomes for hip replacement surgery. A cluster randomised controlled trial (PROFILE) was conducted. Eleven surgeons and 304 patients were randomised to the intervention arm and ten surgeons and 288 patients were randomised to the control arm. Surgeons in the intervention group received peer benchmarked PROMs feedback and an educational session. Surgeons in the control group did not receive feedback or education. The primary outcome measure was the post-operative Oxford Hip Score (OHS). Secondary outcomes were the Hip Osteoarthritis Outcome Score (HOOS), the EQ-5D and the proportion of patient reporting a problem after surgery. Qualitative interviews were undertaken with surgeons in the intervention arm of the trial to examine their views of and reactions to the intervention, and a framework approach was used for analysis.

Results: The quantitative review of 17 studies found weak evidence to suggest that providing PROMs feedback to professionals promotes improvements in patient outcomes. This review identified 16 studies which used PROMs to manage individual patient care and only one study which used PROMs to measure providers' performance. The qualitative review of 16 studies identified the barriers and

facilitators to the use of PROMs in practice. Four major themes emerged: practical considerations, attitudes towards the value of the data, methodological concerns and the impact of feedback on patient care. The PROFILE trial found no significant difference in outcomes between surgeons in the intervention and control arm.

Primary outcome data were available for 11 intervention surgeons with 215 patients and for 10 control surgeons with 217 patients. The mean post-operative OHS for the intervention group was 41.1 (95% CI 40.1-42.0) and for the control group was 41.9 (95% CI 41.0-42.7). The adjusted effect estimate was -0.7 (95% CI -1.9-0.5, P=0.2).

Outcomes for patients in both groups improved over the course of the trial, although the differences between pre- and post-feedback outcomes were not statistically significant. Similar results were found for the secondary outcomes. Interviews with 11 surgeons after they received the intervention revealed mixed opinions about the value of the peer benchmarked PROMs data. Many surgeons appreciated the feedback as it reassured them that their practice was similar to their peers. However, their reluctance to use the information in practice related to conceptual, methodological and practical concerns associated with the collection and use of PROMs data.

Conclusion: This research adds significantly to knowledge in this field as it presents the first randomised controlled trial which examines the impact of providing surgeons with peer benchmarked PROMs feedback and the first qualitative study which explores surgeons' reactions to the feedback. A number of recommendations for the future design of a PROMs feedback intervention emerged from this research. It is important to consult with professionals at the developmental stage of a PROMs feedback initiative, communicate with professionals about the objectives of the data collection, educate professionals on the properties and interpretation of the data, and

support professionals in using the information to identify areas for improvement. It is also imperative that the burden on patients and staff is minimised so that the collection and dissemination of PROMs information integrates more seamlessly into daily working patterns.

1

Introduction & Background

Chapter 1- Introduction & Background

1.1 Introduction

In this thesis, the use of patient-reported outcome measures (PROMs) as a tool to improve the quality of patient care will be examined. PROMs are questionnaires which assess patient's health, health-related quality of life and other health-related constructs (1). In recent years, there has been a growing trend to systematically collect and feedback PROMs information to healthcare professionals (2-3).

This research is presented through a series of papers. First, a background chapter explains the scope of and rationale for the thesis. Second, a systematic review presents the evidence on the impact of providing PROMs feedback to professionals. Third, another systematic review synthesises the qualitative evidence on professionals' experiences of receiving and using PROMs feedback in practice. Fourth, a cluster randomised controlled trial examines the impact of providing consultant orthopaedic surgeons in Ireland with peer benchmarked PROMs feedback. Fifth, a qualitative study explores surgeons' views of and reactions to the PROMs feedback received in the trial. Finally, a discussion chapter combines the evidence from the four studies outlining the implications of the findings on policy, practice and research.

1.2 Background

Quality in healthcare can be defined as 'the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge' (4). A high quality service should be safe, effective, patient-centred, timely, efficient and equitable (4-5). A failure to achieve such desirable features can be attributed to issues related to

underuse, overuse and misuse of resources (6). It is widely accepted that poor quality care is a persistent issue facing healthcare services. Extensive research has demonstrated significant variations in healthcare practices, as well as an unacceptable level of errors (4, 7). Although significant effort has been placed on trying to improve care, the process of maintaining and improving quality is an ongoing challenge due to the unstandardized, non-linear and complex nature of medical decision-making (8).

In an attempt to achieve high quality care, two practices can be adopted: quality assurance and quality improvement. Both of these practices are based on the premise that you cannot manage quality if you do not measure quality. Quality assurance involves monitoring care to ensure that it is delivered in a consistently high quality manner across all professionals (5, 9). Care is usually monitored against standards and if levels of compliance breach a certain threshold, corrective action is taken (5, 10). Quality improvement aims to continuously improve services by reducing variation in practice or shifting quality indicators in a desired direction. The main difference between the two approaches is that quality assurance normally focuses on the performance of those identified as being outliers, while quality improvement focuses on improving care across all providers by continuously looking for a better way to provide care (5). The focus of this thesis is on quality improvement activities. In order to promote real quality improvement, quality does not only have to be measured but great consideration is required to ensure that it is measured accurately. Comparable to a weighing scale, or any good measurement tool, a measure should possess the following characteristics: it should produce similar findings if you measure something numerous times (reliability), it should capture what it intends to

measure (validity), and it should be able to detect significant changes over time (responsiveness).

1.2.1 Quality improvement in healthcare

Initial improvements in the quality of care can be traced back to the 19th century.

Among these pioneering efforts were the promotion of hand washing by Ignaz Semmelweis, the introduction of outcomes research by Florence Nightingale and the establishment of performance monitoring by Ernest Codman (6, 9, 11).

However, the first system-wide approaches to improve the quality of healthcare stem from the United States. Policy makers identified a need to improve care after recognising that the health system was not achieving its purpose – to provide care for the entire population and in particular the aged and disabled. Medicare and Medicaid were established to address this inequity. A set of conditions for hospital participation within these programmes were developed which stipulated, for the first time, that hospitals were required to evaluate how care was being delivered.

Utilisation Review Committees were established to coordinate the reviews (6, 12) and subsequently, Medical Care Review Organisations were developed to monitor the delivery and quality of inpatient and ambulatory care. The success of these schemes led to the development of legislation and the creation of specific organisations to assess the quality of care. For example, Peer Review Organisations were appointed with the authority to implement solutions for given problems by enforcing reviews, further education, disciplinary action and loss of billing privileges (6).

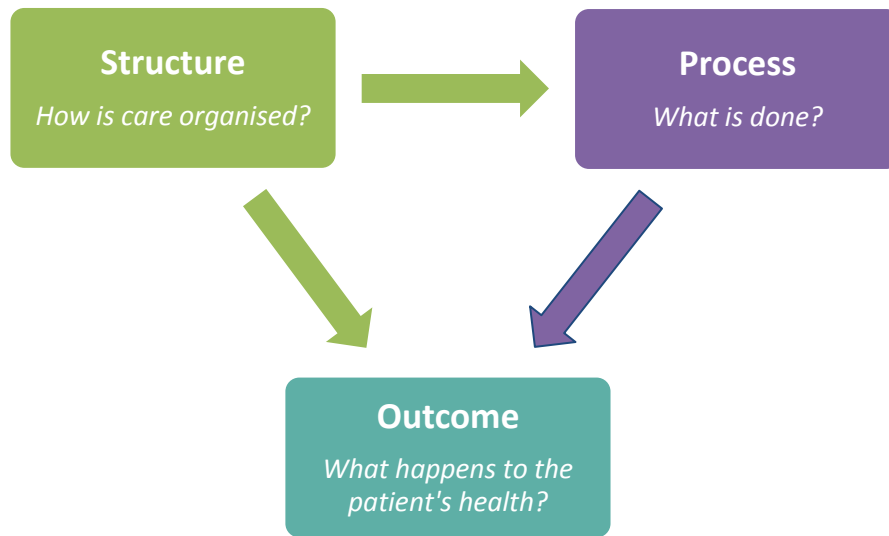
From the 1970s, many State run organisations began to emerge to support quality improvement in healthcare: the Institute of Medicine was developed to evaluate, inform and improve the quality of care (13); the Agency for Healthcare Research and

Quality was created to address geographic variations in practice patterns (14); and the National Committee for Quality Assurance was established to manage accreditation programmes (15). Furthermore, organisations such as the Institute for Healthcare Improvement (16) and the International Society for Quality in Healthcare (17) began to promote shared learning and collaboration across healthcare systems. Internationally, a host of quality improvement organisations now exist. A number of examples are: the Care Quality Commission in England (18), the Canadian Foundation for Healthcare Improvement (CFHI) (19), the Australian Commission on Safety and Quality in Healthcare (20), and the Health Information and Quality Authority in Ireland (21).

1.2.2 Framework for quality improvement

This quality improvement movement has been heavily influenced by a framework developed by Avedis Donabedian. In 1966, Donabedian proposed a model which attributed care structures and processes to patient outcomes (6, 9). Structures focus on healthcare infrastructure and institutional factors including buildings, professionals, regulatory and financing environments. Processes of care include services provided to a patient during their healthcare journey such as getting examinations, tests, and being prescribed medications. Outcomes are the final effect of healthcare interventions on the patient's health and wellbeing (22-23). The objectives of quality improvement interventions can be classified on the basis of this framework (Figure 1).

Figure 1: Structure-process-outcome framework



Ref: (24)

1.2.3 Quality improvement strategies

There is no single definition for quality improvement. For the purpose of this thesis, Dr John Øvretveit's definition will be used which defines quality improvement as 'better patient experience and outcomes achieved through changing provider behaviour and organisation through using a systematic change method and strategies' (25). This definition was selected as it highlights the two key elements of a quality improvement plan. The first involves choosing a specific quality improvement intervention (change method or strategies) and the second involves defining the objective of the strategy (changing provider behaviour and organisation). Each of these elements will be elaborated upon in turn.

Firstly, there are many quality improvement interventions that can be used to promote effective change in care. The Agency for Healthcare Research and Quality drew on previous efforts to categorise quality improvement interventions. They identified nine possible interventions which can be adopted to improve patient care:

provider reminder systems; facilitated relay of clinical data to providers; audit and feedback; provider education; patient education; promotion of self-management; patient reminder systems; organizational change; and financial, regulatory or legislative incentives (26) (Table 1).

Table 1: Categories of quality improvement interventions

Intervention	Description	Evidence of effectiveness
Provider reminder systems	Information provided to prompt a clinician to recall information, to prompt consideration of a specific process of care or to follow evidence-based care recommendations.	Modest effects (27)
Facilitated relay of clinical data to providers	Transfer of clinical information collected from patients and relayed to the provider (i.e., the telephone transmission of a patient's blood pressure measurements). May be some overlap with provider reminder systems.	Mixed evidence (28-29)
Audit and feedback	Summary of clinical performance for healthcare providers or institutions over a specific period of time and reported either publicly or confidentially to the clinician or institution. This includes benchmarking of process or outcomes of care.	Modest effects (30)
Provider education	Educational workshops, meetings, continuing medical education, lectures, and educational outreach visits.	Mixed evidence (31)
Patient education	Patient education, either individually or as part of a group or community through print/audio-visual educational materials.	Mixed evidence (32-34)
Promotion of self-management	Distribution of materials or access to a resource that supports patients to manage their condition, the communication of useful clinical data to patients, or follow-up phone calls from the provider to patients with recommended adjustments to care.	Modest, short-term effects (35)
Patient reminders	Effort directed by providers toward patients to encourage them to keep appointments or adhere to other aspects of self-management.	Modest, short-term effect (36)

Intervention	Description	Evidence of effectiveness
Organizational change	Disease or case management; team or personnel changes; communications, case discussions, and the exchange of treatment information between distant health professionals; Total Quality Management or Continuous Quality Improvement techniques for measuring quality problems and changes in medical records systems.	Mixed evidence (37-40)
Financial, regulatory or legislative incentives	Positive or negative financial incentives directed at providers or patients; system-wide changes in reimbursement; changes to provider license requirements; and changes to institutional accreditation requirements.	Scant evidence (41-42)

REF: (26)

Currently, there is little consensus about which quality improvement intervention is most effective (30, 43-47). For example, reminder systems can be effective in improving processes of care particularly for prescribing practices (27), provider education can stimulate modest improvements to professional practice (31, 46), patient education can promote small short-term improvements in self-management strategies (35), clinical decision support systems can improve preventive care and prescription practices (48), audit and feedback can lead to small but potentially important improvements in professional practice (30), and the evidence on financial or regulatory incentives is too weak to draw definitive conclusions about effectiveness (41-42). These findings tend to echo the conclusion a review of interventions to improve professional practice which stated that there are no ‘magic bullets’ for improving quality but appropriately designing and implementing interventions can lead to important changes in care (47).

Secondly, the objectives of the strategy can be linked to Donabedian’s structure-process-outcome framework (49-51). Quality improvement strategies tend to target

changes in care processes and outcomes over structures, as often structures of care are less amenable to change (47).

The Cochrane Effective Practice and Organisation of Care Review Group categorise the objectives of a quality improvement strategy as a change in the following: clinical prevention services, diagnosis, test ordering, referrals, procedures, prescribing, general management of a problem, patient education/advice, professional-patient communication, record keeping, resource use, discharge planning and patient outcomes (52). The first 12 objectives on the list outline specific processes of care that can be targeted to improve quality. The final objective 'patient outcomes' refers more generally to efforts which promote improvements in outcomes. An outcome includes measures of health status, morbidity and mortality (53) (refer to Figure 1).

This thesis focuses on the use of audit and feedback as a quality improvement intervention. The use of audit and feedback is unique in that the objective can be to support changes in processes of care (e.g. to improve prescribing) or outcomes of care (e.g. reduce mortality). The following section will discuss audit and feedback in more detail and will elaborate further on the difference between focusing on processes versus outcomes, as the mechanisms by which change may occur are conceptually very different (49).

1.2.4 Audit and feedback

Audit and feedback is defined as 'summaries of clinical performance (audit) over a specified period of time, and the provision of that summary (feedback) to individual practitioners, teams or healthcare organisations' (49). Individual practitioners, teams or healthcare organisations will be referred to as 'providers' for the remainder of this thesis. The rationale for using audit and feedback to promote quality improvements

is based on a number of assumptions: healthcare providers have a limited ability to accurately self-assess practice, they have an inherent motivation to improve care and lastly they are unaware of their relative performance (54). They may be prompted to change practice or the organisation of care if the feedback highlights that their current practice or patient outcomes are inconsistent with peers (30).

It is now accepted that audit and feedback can be effective in improving care (30).

Audit and feedback is a highly variable intervention with many components. These include the method of auditing (chart review, computerised records, observation, questionnaire); the level of aggregation of the data (patient, physician, hospital, trust); the setting (community, hospital, specialist); the professional (trainee, physician, specialist, non-medical); the patient population under study (general patient, specific patients focusing on disease or particular characteristics); and the comparison group (time period, national average, benchmarking providers, population norms). In addition, the presentation of the feedback can differ in content, timing, intensity and format (55). Previous research has identified that the impact of audit and feedback tends to be greater when feedback is provided frequently (56), in writing with specific suggestions (57) and is generated from a reliable source (58). Intuitively, the impact of the intervention is larger when baseline adherence to recommendations is low (56).

As mentioned earlier, the objective of an audit and feedback intervention can be to support changes in processes of care by offering providers information about 'what they do' or outcomes of care by offering providers information about 'what results they achieve'. The implementation pathway when feeding back information on processes is more straightforward than when feeding back information on outcomes

as the target of change is explicit. Two examples will be provided to clarify this last point.

Feedback based on processes of care presents information to providers about current clinical practice (59). Process measures are direct indicators of quality and are easy to interpret (59-60). An example is: undertaking an audit of aspirin prescribing for patients with heart disease and feeding back the information about differences in practices across general practitioners (61). In this situation, the target of change is obvious— change prescribing patterns by prescribing aspirin to appropriate patients (Figure 2). Much of the literature on the impact of audit and feedback focuses on prescribing and test utilisation (30). The possible reasons why the research to date tends to focus on these areas are: firstly, the feedback promotes a change in a specific task (e.g. prescribing) so the intervention is relatively simple to implement; and secondly, it makes sense to measure processes of care in a situation where best practice guidelines exist as the extent to which providers have adopted these practices can easily be established (62-63).

Figure 2: Implementation pathway – audit and feedback of a process measure



Feedback based on outcomes of care presents information to providers about current performance in terms of patient outcomes. Outcome measures are intrinsically valuable as they are an overall indicator of care (59-60). However, the implementation pathway is more complex when feeding back outcomes as the target of change is not explicit. An example is: undertaking an audit of mortality rates for

coronary artery bypass grafts and feeding back information about differences in outcomes to surgeons (64). In this situation, the change required to improve outcomes is not obvious—the feedback identifies that variations in outcomes exist but providers are required to identify what they could do differently to improve patient outcomes (57, 60) (Figure 3). This may involve having to undertake additional audit or research before the solution to the problem is identified and the change in practice is implemented. Thus, this approach prompts providers to identify which processes of care they need to change if they are not performing as well as their peers (59-60). The difference between Figure 2 and Figure 3 is that the target of change is defined in Figure 2 but not defined in Figure 3.

Figure 3: Implementation pathway- audit and feedback of an outcome measure



In recent years, there has been a growing interest in focusing on improving outcomes of care. Outcome measures have an inherent value, particularly when clear guidelines on the most effective course of treatment do not exist (3, 65-66). This thesis focuses on the application of audit and feedback and outcomes measurement as this approach is becoming more popular, but the evidence on the effectiveness of feeding back outcome measures to providers remains unresolved (3, 67-68). The following section will provide an overview of the ‘outcomes movement’.

1.2.5 Outcome measurement

The United States led the shift in focus from process measurement to outcome measurement in the 1980s. This change in emphasis is commonly referred to as ‘the outcomes movement’ (69-70). Donebedian argued that ‘Outcomes, by and large, remain the ultimate validators of the effectiveness and quality of medical care’ (71). The focus on measuring outcomes is driven by a belief that this information has the potential to streamline how healthcare is organised and delivered (72).

Traditional outcome measures focused on mortality and morbidity such as death and complications after surgery. While these are extremely important indicators of care, they have limitations for quality improvement purposes. For example, these indicators tell us little about the quality of care received for the vast majority of patients who undergo non-surgical or lower risk healthcare interventions (73). The importance of expanding outcome measurement beyond morbidity and mortality has become increasingly pronounced, particularly with the rise in chronic conditions. In which case, healthcare interventions aim to improve the physical, psychological and social aspects of life (71, 74). To address this need, the concept of quality of life measurement emerged as a popular solution. Quality of life measurement is linked to the belief that healthcare should move from a biomedical to biopsychosocial approach to managing care (73).

One of the first quality of life instruments was the Karnofsky Performance Scale. This is a simple scale ranging from 0 (dead) to 100 (no evidence of disease) which was designed to be completed by the provider on behalf of the patient (75). Over the past three decades, quality of life measures have evolved into many health constructs including health-related quality of life, health status, functional status and emotional well-being. Recently more emphasis has been placed on acquiring this information

from patients themselves (22). The interest in capturing outcomes reported by patients is driven by the perception that patients are best positioned to judge their own welfare (71). The term PROM has been adopted to describe any outcome data provided by the patient (76-77). Nowadays, there are hundreds of PROMs which have been categorised into disease-specific, site or region-specific, dimension-specific, generic, summary items, individualised and utility measures (76).

The value of using PROMs was first realised for evaluation purposes in clinical research (78-79). Subsequently, the potential of linking audit and feedback with the collection of PROMs to promote quality improvement emerged with two distinct goals: to guide individual patient management and to measure the performance of providers. The following section will elaborate on the different ways PROMs can be used to promote quality improvements in care.

1.2.6 Different uses of PROMs

The value of PROMs first emerged for research purposes—testing the effectiveness of different treatments in clinical trials— as biomedical measures alone may fail to capture issues that are important to patients (78-79). Subsequently, the Food and Drug Administration (FDA) endorsed the use of PROMs to identify treatment effects of medicines and to examine the effectiveness of treatments from the patients' perspective (80). Furthermore, the National Institute for Clinical Excellence (NICE) also decided to routinely use PROMs to appraise health technologies (3).

In turn, the value of using PROMs to capture information for quality improvement purposes was recognized. This particular function involves collecting PROMs and providing healthcare providers with feedback, as a belief emerged that this additional patient focused information would promote changes to enhance care (81-82).

However, the mechanisms by which changes in care may occur differ depending on

the level at which the data is fed back to the provider. Feedback can be delivered at the individual patient level or the aggregated level.

The use of PROMs as a quality improvement tool was initially employed to guide individual patient management. In this context, PROMs are routinely collected from patients and the individual level information is fed back to professionals usually during a consultation. If measures are adequately developed, PROMs feedback about individual patients can help to screen for undiagnosed problems, to assist in identifying and prioritising health concerns, or to promote patient–physician communication (81-82). The feedback is thought to stimulate the professional to manage the patient’s care differently such as changing medicines, ordering further tests, referring the patient to other healthcare professionals, or by encouraging the professional to advise and educate the patient on effective management of their problem. Furthermore, improved communication between the patient and professional is thought to lead to a greater understanding of complex personal factors, and to encourage the professional and patient to set shared goals for treatment. This may lead to better targeting of treatment towards issues that are important to the patient, in turn promoting better adherence to treatment, greater satisfaction and ultimately better health outcomes for the patient (81, 83).

Feedback can also be provided for groups of patients by aggregating data to evaluate the performance of healthcare providers, and hence this function serves as a measure of performance (84-85). When the aim of the quality improvement strategy is to improve performance; PROMs are collected from patients, aggregated to the level of the provider and fed back in the form of peer benchmarked reports. One of the first attempts to use PROMs to measure provider’s performance was undertaken in the UK in 1998 by a private healthcare company, BUPA (86). This initiative involved

collecting PROMs data before and after surgery for a range of elective procedures such as hip and knee replacement surgery. The information was aggregated to the level of the hospital and the surgeon, and feedback was provided at regular intervals to the different stakeholders. Peer benchmarking using PROMs was implemented to stimulate continuous quality improvements and to support clinical governance (86). The novel use of PROMs to assess the performance of providers by the private hospital sector inspired the National Health Service (NHS) in England to implement a similar programme.

The NHS PROMs Programme is an example of the most advanced PROMs initiatives (3). For this reason, this thesis focuses on the methodology used by NHS PROMs Programme. The following section will provide an overview of the Programme and describe the journey of PROMs into policy in England, before outlining the mechanisms by which this programme is expected to promote quality improvements.

1.2.7 The NHS PROMs Programme

In 2009, the NHS PROMs Programme was introduced in England. This programme instituted the collection of PROMs as a mandatory requirement for audit for four common surgical procedures: hip replacement, knee replacement, hernia repair and varicose vein surgery. Similar to the method employed by BUPA, questionnaires are collected from all eligible patients prior to surgery and either three (varicose vein surgery) or six months after surgery (hip replacement, knee replacement and hernia repair). The patient-provided information is published online by the Health and Social Care Information Centre at the level of the Trust. Analysis of pre-operative data was released in April 2010, analysis of post-operative data was released in

September 2010 and from August 2011 the PROMs data has been published by year to enable comparisons (87).

The journey of PROMs into policy

The first reference to the use of PROMs as a tool to measure performance in the NHS was published in 2004 by Dr. Foster Ethics Committee (2). This report outlined the benefits of routinely collecting PROMs to inform the revalidation of clinicians' licences, to manage performance of hospitals and to guide patient choice. In the following years, researchers in the London School of Hygiene and Tropical Medicine were contracted by the Department of Health to identify the most appropriate measures to use, and to determine the feasibility of collecting and using PROMs in practice for elective surgery (1, 88). In 2007, Lord Darzi recommended in the NHS Next Stage Review Interim Report that PROMs should be used to drive quality improvements across the NHS (89) and later that year the NHS announced that from April 2009 providers would be obliged to collect PROMs from all NHS patients undergoing one of the four elective procedures included in the Programme (90). In 2008, the Government published the NHS Next Stage Review which detailed their intention to link the PROMs data to providers' payments (91). Prior to the introduction of the PROMs programme in 2009, the Department of Health published a guidance document on the routine collection of PROMs (92). A second independent report was commissioned by The King's Fund in 2010 to outline how PROMs could be used most effectively to drive quality improvements in the NHS (3). Later that year, in a white paper the Government stated that they would expand the PROMs programme beyond surgical procedures (93). From 2010, a series of publications by the NHS Quality and Outcomes Framework detailed the inclusion of PROMs as indicators within the broader quality agenda and in 2013, the framework

highlighted that the London School of Hygiene and Tropical Medicine were investigating the potential of routinely collecting PROMs for dementia care (94-97). Finally, in a consultation document published last year, the NHS released details on their first attempt to link providers' payments to the PROMs Programme based on participation rates of Trusts (98) (Table 2).

Table 2: The journey of PROMs into policy in England

Reference	Agency & type of report	Title	Aim	Significance
Appleby, 2004 (2)	Dr. Foster Ethics Committee Commissioned report	Measuring success in the NHS- using patient assessed health outcomes to manage the performance of healthcare providers	To outline the benefit of routinely collecting PROMs to monitor and manage the performance of providers as a means of facilitating a system-wide refocus of the NHS on health	<i>Details the mechanisms by which PROMs may stimulate change</i> Direct approaches: planned actions in response to evidence by management or regulators to reward or penalise providers based on performance. E.g. licencing, financial rewards, contractual arrangements and disciplinary actions Indirect approaches: behavioural responses to pressures linked to evidence on performance. E.g. greater clinical governance, patient and commissioner choice
Smith, 2005 (88)	London School of Hygiene and Tropical Medicine Literature Review	Patient-Reported Outcome Measures (PROMs) for routine use in Treatment Centres: recommendations based on a review of the scientific	To present a review of the literature to identify disease (or procedure) specific PROMs in five areas of surgery (cataract surgery, varicose vein procedures, hip replacement, knee	<i>Outlines the most appropriate measures for five elective procedures</i>

Reference	Agency & type of report	Title	Aim	Significance
		evidence	replacement and hernia repair), generic measures applicable to all surgical areas and also instruments that assess post-operative complications	
Browne, 2007 (1)	Royal College of Surgeons England and London School of Hygiene and Tropical Medicine Pilot study	Patient Reported Outcome Measures (PROMs) in Elective Surgery	To determine the feasibility of collecting pre- and post-operative PROMs from patients undergoing elective surgery and to investigate how such data could best be analysed and presented	<i>Outlines the feasibility of collecting and using PROMs in practice</i>

Reference	Agency & type of report	Title	Aim	Significance
Lord Darzi, 2007 (89)	Department of Health Interim report	Our NHS Our Future - NHS Next Stage Review Interim Report	To outline the vision of the NHS, ahead of the publication of the next stage review	<i>Recommends the use of PROMs by the NHS</i> ‘...build on recent advances in measuring outcomes as assessed by patients themselves, and make these patient-reported outcome measures a stronger part of our approach to clinical quality’
Department of Health, 2007 (90)	Department of Health Policy document	Guidance on the Standard NHS Contract for Acute Hospital Services	To provide guidance for NHS commissioners and service providers in England on the standard contract introduced from April 2008	<i>Stipulates the mandatory collection of PROMs within the NHS</i> ‘The contract supports an increasing emphasis on commissioning for outcomes by introducing a new requirement, in Schedule 5, to report from April 2009 on patient-reported outcome measures (PROMs). These will cover NHS patients undergoing hip and knee replacements, groin hernia repair and varicose vein ligation’
Lord Darzi, 2008 (91)	Department of Health Command paper	High quality care for all: NHS next stage review	To outline the vision for the future of health and healthcare	<i>Recommends linking PROMs to payments and expanding the use of PROMs to promote effectiveness of care</i> ‘...make payments to hospitals conditional on the quality of care given to patients as well as the volume. A range of quality measures covering ...and patient’s views about the success of their treatment (known as patient-reported outcome measures or

Reference	Agency & type of report	Title	Aim	Significance
				PROMs) will be used' 'Understanding success rates...from the patient's own perspective which will be measured through patient-reported outcomes measures (PROMs)'
Department of Health, 2009 (92)	Department of Health Guidance document	Guidance on the routine collection of Patient Reported Outcome Measures (PROMs)	To provide guidance on the routine collection of PROMs for elective procedures from 1 April 2009	<i>Outlines the implementation of PROMs and the specific value of using PROMs as a quality improvement tool</i> -Evaluate clinical quality: PROMs data can be used by clinicians, managers, regulators and primary care trust commissioners to benchmark providers' performance. PROMs can also be used to guide clinical audit and to inform patients and GPs choice -Research what works: efficacy and cost-effectiveness of different technical approaches to care can be evaluated using PROMs in association with other measures - Assess the appropriateness of referrals to secondary care -Support the reduction of inequalities -Empower commissioners to establish the quality of services for which they contracting with providers

Reference	Agency & type of report	Title	Aim	Significance
Devlin, 2010 (3)	The King's Fund and the Department of Health PROMs Stakeholder group Commissioned report	Getting the most out of PROMs- putting health outcomes at the heart of NHS decision-making	To provoke and encourage thinking about the wide range of ways in which PROMs data can be used to inform decisions in the NHS	<i>Highlights how PROMs could be used most effectively to drive quality improvements.</i> -Patient choice: guiding patient decisions on where to receive care thereby stimulating provider response to enhance reputation and achieve clinical improvement -Commissioner choice: guiding GPs and commissioner's decisions on obtaining and purchasing services on behalf of patients, and ensuring value-for-money -Managing clinical quality: creating dialogue between managers and clinicians to identify actions to improve quality and efficiency, linking to Hospital Episode Statistics to examine reasons for variations, publically reporting to benchmark performance and improve accountability, and introducing payment for performance -Clinical decision making: identify benefits of treatment, stimulating joint-decision making and informing referral management

Reference	Agency & type of report	Title	Aim	Significance
Department of Health, 2010 (93)	Department of Health White paper	Equity and excellence: Liberating the HNS	To outline the Government's long-term vision for the future of the NHS: put patients at the heart of everything the NHS does; focus on continually improving those things that really matter to patients - the outcome of their healthcare; empower and liberate clinicians to innovate, with the freedom to focus on improving healthcare services	<i>Endorses the extended use of PROMs</i> 'Information generated by patients themselves will be critical to this process, and will include much wider use of effective tools like Patient-Reported Outcome Measures (PROMS), patient experience data, and real-time feedback. At present, PROMs, other outcome measures, patient experience surveys and national clinical audit are not used widely enough. We will expand their validity, collection and use. The Department will extend national clinical audit to support clinicians across a much wider range of treatments and conditions, and it will extend PROMs across the NHS wherever practicable'

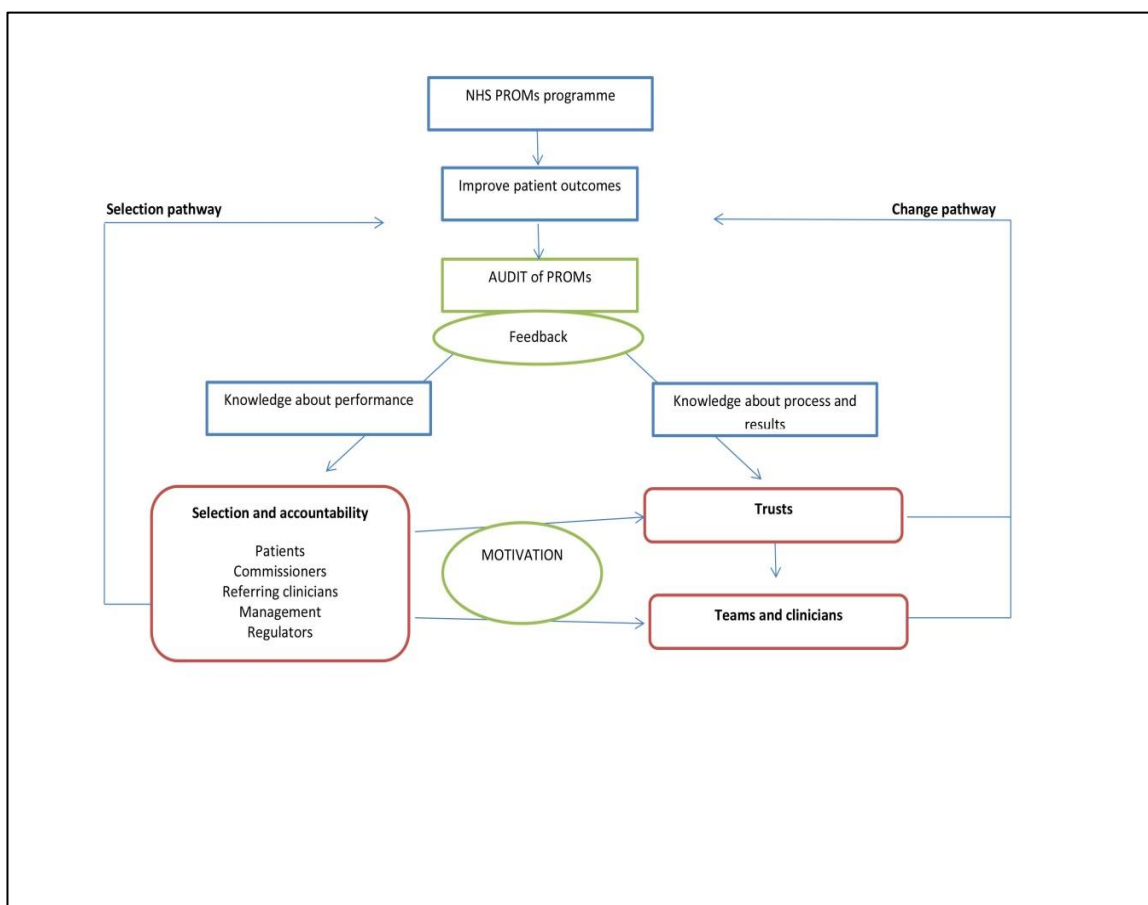
Reference	Agency & type of report	Title	Aim	Significance
Department of Health, 2010 (94)	Department of Health Consultation/ Discussion document	Transparency in outcomes: a framework for the NHS	To introduce the NHS Outcomes Framework	<i>Outlines the integration of PROMs into the Outcomes Framework</i> ‘The indicators included in the framework therefore need to cover both clinical outcome measures as well as patient reported outcome measures (PROMs)’
Department of Health, 2010 (95)	Department of Health Policy document	The NHS Outcomes Framework 2011/12	To outline the NHS Outcomes Framework. It builds on the proposals published for consultation in Transparency in outcomes – a framework for the NHS and the responses received to that consultation	<i>Details an intention to expand the use of PROMs within the Outcomes Framework</i> ‘PROMs currently exist for four elective procedures. They are included in this first framework, with a view to considering development of further PROMs in light of the NHS Outcomes Framework’

Reference	Agency & type of report	Title	Aim	Significance
Department of Health, 2013 (99)	Department of Health Policy document	The NHS Outcomes Framework 2014/15	To provide an update on the progress that has been made to develop existing indicators in the NHS Outcomes Framework	<i>Details an intention to expand the use of PROMs within the Outcomes Framework</i> ‘The Department of Health has commissioned a research team at the London School of Hygiene and Tropical Medicine to investigate the potential for a routine Patient-Reported Outcome Measure for dementia, including where necessary a measure for completion by a relevant person other than the patient’
NHS, 2013 (98)	National Health Service Consultation document	2014/15 National Tariff Payment System: A Consultation Notice	To outline the proposed national tariff as required by section 118 of the 2012 Act	<i>Announces the intention to link PROMs participation rates to payments</i> ‘Collecting data on quality of care through PROMs and clinical audits is important as these data underpin high quality care and can inform choices made by commissioners and patients, as well as the development of policy. By linking payment for the BPT (Best Practice Tariff) to achieving minimum levels of compliance and consent rates, we aim to improve data collection, submission and response rates’ ‘...a minimum PROMs participation rate of 50%’

Quality improvement pathways

The mechanism by which the NHS PROMs Programme may drive quality is implicit, rather than explicit, within the documents published by the Department of Health. As such, one can only speculate about how the use of PROMs as a performance measurement tool will promote improvements. Berwick identified that improvements in quality can be promoted through two pathways- a selection pathway and a change pathway (100) (Figure 4).

Figure 4: Quality improvement pathways



Ref: (100)

In the context of the NHS PROMs Programme, the model suggests that a selection pathway would be facilitated through increased accountability by offering patients, commissioners, referring clinicians, management, and regulators with the evidence to select providers based on their performance. Selection may occur as a result of a

change in demand for the service or through a quality control process undertaken by management or regulators. Firstly, patients, commissioners and referring clinicians naturally demand services from the best performers thereby reducing or eliminating a demand for services provided by the worst performers. Secondly, regulators may revalidate licences to practice and management may allocate workloads based on the PROMs data, thereby restricting the worst providers from practicing. In this way, the selection pathway therefore either ‘culls’ poorer performers from providing services or forces providers to enhance quality. The threat of selection should motivate providers to establish what changes are required to improve performance.

The model suggests that a change pathway would be promoted by increasing awareness of one’s performance relative to their peers. The process of performance measurement is linked to a number of strategies that are thought to motivate professionals to improve. For example, monitoring alone is believed to influence performance through a psychological force known as the Hawthorne effect, which generates a heightened self-awareness of ones actions and the consequences of those actions (101); benchmarking is based on the premise that professionals have an intrinsic competitive nature and that peer comparison provides the necessary motivation to stimulate change, by reminding or forcing individuals into action (102); and lastly incentivising providers by linking payment to performance rewards providers who continuously find a better way to deliver care. Change can be promoted at the hospital, team or individual provider level. For example, a hospital may organise a peer mentoring programme where poor performers learn from the best performers, a team may undertake an audit of complications after surgery and implement greater hygiene practices, and an individual clinician may decide to change his/her technique to a less invasive surgical approach.

Summary of what is known and not known

It is clear that the quality of healthcare is a persistent problem facing policy makers. One popular intervention that can be used to improve quality is audit and feedback. In recent years, there has been an increasing interest in focusing on outcome measurement and in particular on PROMs. The NHS PROMs Programme is a good example of a quality improvement strategy which incorporates the use of audit and feedback and PROMs. It is clear that the Department of Health intends to push forward with the PROMs agenda in England, and interest in the use of PROMs as a quality improvement tool is gaining momentum internationally (3, 72-73, 103-108). However, to date there has been little effort placed on trying to establish if and how PROMs feedback may lead to improvements in patient outcomes. The impact of using PROMs to improve the quality of care has been documented through a number of systematic reviews. However, there is weak evidence to suggest that the feedback improves patient outcomes (83, 109-112). Therefore, this thesis seeks to establish the impact of feeding back PROMs information to providers on improving patient outcomes and to examine providers' experiences when using PROMs as quality improvement tools.

1.2.8 Context of the research

It is important to outline the context in which a quality improvement intervention is implemented because factors such as the organisation, funding, structure and culture of a healthcare system can influence findings. This section will outline the context of this research in Ireland.

Hip replacement surgery

Hip replacement surgery is one of the four elective procedures selected for the NHS PROMs Programme (3). Hip replacement surgery provides a long-term solution for worn or damaged hip joints causing pain and limited mobility. The operation replaces the natural socket and the rounded ball at the head of the thigh bone with artificial parts. The procedure is associated with improved physical function and reduced pain (113).

Evidence suggests that the practice of hip replacement surgery varies considerably between providers (114). There is much heterogeneity in surgical technique as there are over 100 varieties of hip prostheses, multiple bearing couples and several surgical approaches (115). Furthermore, practices vary between settings which support different patient pathways and governance structures (116). The recall of the DePuy Articular Surface Replacement in 2010 emphasised the need to assess and manage the delivery of care for hip replacement surgery (117).

Hip replacement surgery was chosen as the focus for this thesis as high volumes of this procedure are performed yearly in Ireland. A procedure with high volumes was necessary for this study both to accurately benchmark providers' performance and to undertake a cluster randomised controlled trial within a sensible timeframe. The number of hip replacements undertaken in Ireland in 2011 was 2997 and in 2012 was 3,132 (118-119). Hip replacement surgery is performed in 12 public hospitals and 14 private hospitals with elective orthopaedic units. Currently, there are 123 consultant orthopaedic surgeons in Ireland: 87 work in public hospitals and also undertake some private practice, and 36 work in private practice only. Approximately 66 of the consultants perform hip replacement surgery, and this number has increased in recent years. The decision to commence a national joint registry in Ireland was granted by

the Department of Health in 2011 (120). The registry is currently being developed by the National Office of Clinical Audit (120).

Organisation of the health service

The population of Ireland is currently 4.58 million. The population is increasing and aging which is putting a greater demand on health services (121).

The Irish healthcare system has undergone two major reforms in the past 40 years and is currently undergoing a third (122). The first occurred in 1970 when the management of services were removed from local authorities and re-organised into eight regional health boards. In the 1990s, one of the boards was subsequently divided into three smaller boards. The second reform occurred in 2004 in response to the publication of the Prospectus report. The aim of the reform was to make health services more unified, efficient and less vulnerable to local and parochial pressures (123). The Prospectus report recommended a reorganisation which merged the 11 boards into a national health service called the Health Service Executive (HSE). The Department of Health handed the duty of executing policy, administration and management to the HSE, and the Minister for Health and Children held overall responsibility for the Executive. As part of this reform, the Health Information and Quality Authority (HIQA) was established in 2007. HIQA is currently responsible for providing health information, setting and monitoring standards, promoting and implementing quality assurance programmes and overseeing health technology assessment (123-125). However, the operational performance and outcomes of the HSE have been strongly criticised by the public. The major grievances with the delivery of care are long waiting times in the Emergency Department (ED), lengthy waiting lists for individual procedures and treatments, and the public-private funding model which has led to the creation of a two-tier system. The third reform was

announced in 2012 in the white paper- Future Health: A Strategic Framework for Reform of the Health Service 2012-2015 (122). This reform aims to tackle the inequity in the system by removing the two-tier system and improving performance by reorganising the financing and structure of the service.

Funding of the health service

In Ireland, the health system is predominantly tax funded. Tax contributes approximately 67 per cent of the total health care expenditure. The remainder of the funding is made up of out-of-pocket payments (fees paid to GPs, consultations in private practice, physiotherapists, dentists, opticians and charges for medicines) and private health insurance (124). Approximately 40% of the population have medical cards which entitle holders to most services free of charge. About 43% of the population have voluntary private health insurance which entitles holders to greater access to secondary care services. Finally, almost 25% of the population have neither medical insurance nor a medical card and require some out-of-pocket payments for both primary and secondary care services. Some people have both a medical card and private health insurance (125).

Hospitals receive a fixed budget each year (124). This funding model does not incentivise efficiencies within the system as there is no reward for increasing throughput of public patients. This has resulted in the development of the two-tier system favouring patients that have insurance (125). However, there is increasing unrest regarding the inequity within the system (122). The current Government plans to eliminate the mixed financing system by introducing universal health insurance. This new model will be founded on principles of social solidarity where everyone will be insured for a standard package of primary and secondary care services. Those that cannot afford the mandatory health insurance will be supported in paying the

cost of the insurance. It has been recommended that insurance will be provided under a multi-payer insurer model (122). This will tie in with a new financing system called ‘Money Follows The Patient’ which will ensure that hospitals are paid according to productivity (126).

Primary care

The public’s main access point to the health service is through primary care or General Practitioners (GPs). GPs are self-employed in Ireland working within approximately 1,600 practices nationwide (123). In 2001, the government attempted to develop primary care by proposing a wider availability of services through the establishment of multi-disciplinary primary care teams and co-operatives. However, this policy was not adequately rolled-out across the country and today primary care remains fragmented and under-resourced (127).

GPs are regarded as the “gatekeepers” to secondary care. They are the first point of contact for the patient and a referral letter is required to get access to secondary care. For those with medical cards, access to primary care is free at the point of entry and those who do not qualify for free primary care must pay fees which can vary from €45-70 per visit (123, 125). The current reform also proposes that GP provision will be free at the point of access for all patients (122).

Secondary care

The structure of secondary care is quite unique in Ireland as there are three categories of hospitals: public, voluntary and private. These are all funded and managed in different ways. Public hospitals are managed by the HSE and are state owned. Voluntary hospitals which were initially set up by religious orders are now primarily financed by the State, though these may be still owned and operated by

religious or lay governing boards. The beds within voluntary hospitals may be designated for either public or private use (128). The remainder of hospitals are operated by private insurance companies (128). Hospital consultants are generally employed by the public sector but also work for the private sector. This has had implications on service provision and inequity in the system. While long waiting times are common for appointments in the public system, those with private health insurance can gain faster access to secondary care (125, 128).

As part of the current reform, the government has designed a plan to organise acute services into hospital groups with the intention of establishing independent hospital Trusts. A commissioned report published in 2013 recommended that acute hospitals should be organised into six hospital groups (126). The new configuration of hospital groups plans to reform management teams of voluntary and state owned hospitals within an overall agreed framework for the group (126). This reconfiguration is currently being implemented nationally. Chairs of each group have been appointed, CEOs are currently being recruited and a number of hospital groups have begun organising services accordingly.

Culture of quality improvement

Traditionally, there was an explicit hierarchical structure within the Irish healthcare system. Doctors held an elite position in society which created a culture of obedience, respect and acceptance towards their clinical judgment. Patients were not encouraged to actively participate in care. Audit was not a feature of clinical practice and a lack of transparency was firmly embedded within the system (125, 129). A few high profile examples of unsafe care eventually forced a change of attitudes towards clinical practice in Ireland (129-131). It has taken many years for quality

improvement practices to become established. However, to date many providers still do not have the resources in place to adequately audit their practice (132).

The first attempt to enforce audit into practice occurred in 1997 when consultants' contracts stipulated that clinical audit was a condition of their employment.

However, very little guidance or support accompanied this demand so the adoption of this process in Ireland has been slow (133). In the past number of years, there has been an effort to increase accountability within the system. The Irish Medical Council outlined that consultants are responsible and accountable for their own professional competence and from 2012 consultants are required to demonstrate professional competence by undertaking one relevant audit per year (134). Also in 2012, the HSE undertook a consultation exercise to develop a guidance document for those wishing to carry out clinical audit (132). In addition, the National Office of Clinical Audit was established in 2012. This body aims to design, develop and implement national clinical audit programmes (120). However, the Minister for Health has identified that in order to support real quality improvements within the system, there is a significant need to develop national information systems to facilitate measuring and reporting (122).

In 2011, the Minister tasked the Special Delivery Unit with implementing performance improvements within hospitals focusing on reducing waiting times for emergency care, in-patient and day case procedures, and out-patient visits (135). In 2013, Compstat was introduced as the performance management system aiming to facilitate the management of hospital performance using a monthly scorecard performance report on a suite of metrics (136). Furthermore, a new patient safety agency is to be established, the objective of which is to drive the safety and quality agenda (122). Finally, the new hospital groups have been linked to academic partners

with the purpose of developing a stronger culture of learning, research and innovation (126).

However, amidst these plans is the reality that the health service in Ireland has not yet progressed beyond evaluating rudimentary process measures. In a recent HSE performance report, the areas of focus included: hospital activity, elective inpatients, waiting lists, ED new attendances, ED patient experience times, day care attendances, inpatient discharges, outpatient waiting list, inpatient admission source, emergency response times, emergency admissions, ED trolley performance and access to palliative care (121).

Medical training

A medical degree in Ireland involves a five to six year programme which can be undertaken in one of six medical schools: Trinity College Dublin, the Royal College of Surgeons in Ireland (RCSI), University College Dublin, National University of Ireland Galway, University of Limerick and University College Cork. After graduating, a doctor spends 12 months training as an intern. In this time, they experience a variety of medical specialties to help them decide on which area of medicine they wish to build their career upon. The next stage of training is Basic Specialist Training (BST). There are 10 BST programmes in Ireland: anaesthesia, emergency medicine, general internal medicine, general practice, histopathology, obstetrics and gynaecology, ophthalmology, paediatrics, psychiatry, and surgery (137).

Specialty training programmes in Trauma and Orthopaedic Surgery is delivered by the RCSI. Traditionally, it has taken graduates 13-15 years to train as a consultant orthopaedic surgeon in Ireland. In 2013, the RCSI introduced a new surgical programme which aimed to reduce training to 8 years. After internship, trainees

undergo two years of core surgical training in which time they work under the supervision of a more experienced doctor as a Senior House Officer (SHO). As part of their BST, SHOs are required to pass postgraduate exams as well as completing research and clinical audit projects. To continue training at the registrar level, doctors undergo a competitive process to get accepted into year three of the Higher Surgical Training programme. This involves a further 5-6 years of training while working as a Specialist Registrar. For the final year, doctors are strongly encouraged to undergo sub-speciality training in a centre of excellence either in Ireland or internationally. After successfully completing the programme, the RCSI issue a Certificate of Completion of Surgical Training and the surgeon is entered onto the specialist register. Once on the specialist registrar, surgeons can apply for consultant orthopaedic positions (137-138).

1.3 Overview of the thesis

There is a growing interest internationally in the use of PROMs as quality improvement tools (3, 73). However, there is no strong evidence supporting their use in this context and there is a lack of understanding regarding how the provision of such information may lead to improvements in care (83). This research will employ the methodology used by the largest PROMs initiative— the NHS PROMs Programme.

1.3.1 Research aim, hypothesis and objectives

The **aim** of this thesis is to investigate the value of using PROMs as a quality improvement tool using mixed methods research.

The **hypothesis** of this study is that providing peer benchmarked PROMs feedback to orthopaedic surgeons will stimulate improvements in outcomes for patients undergoing hip replacement surgery.

The research **objectives** of this thesis are as follows:

- To undertake a systematic review to establish the impact of feeding back PROMs information to providers on patient outcomes (Chapter 2)
- To undertake a systematic review to explore the experiences of professionals when using PROMs as quality improvement tools (Chapter 3)
- To undertake a cluster randomised controlled trial to investigate the impact of providing peer benchmarked PROMs feedback to orthopaedic surgeons (Chapter 4)
- To examine the reactions of surgeons to the peer benchmarked PROMs feedback provided in the trial (Chapter 5)
- To collate the evidence and discuss the implications of the findings on policy, practice and further research (Chapter 6)

1.3.2 Research methods

Mixed methods research involves using both qualitative and quantitative approaches in the methodology of a study (139). It can be defined as ‘Mixed methods research is the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g., use of qualitative and quantitative research approaches to data collection, analysis, inference techniques) for the purpose of breadth and depth of understanding and corroboration’ (139). A mixed methods approach is based on the principle that the mixture of both

qualitative and quantitative methods provides a more comprehensive understanding of the research problem than either approach would alone (140).

This mixed method study involves an embedded design; the qualitative research is embedded within the quantitative research. The collection and analysis of both quantitative and qualitative data follows the traditional designs and the ‘mixing’ occurs during the interpretation stage of this thesis (Chapter 6-Discussion). The data for each approach is collected, analysed and reported separately. Interpreting both data sources plays an important role in the overall design, informing the wider research question on the usefulness of PROMs as a quality improvement tool (139-140).

1.3.3 Clarification of the researcher’s role

I, Maria Boyce, believe in using multiple methods of research as long as they are well executed. I am therefore pragmatic in my design and implementation of research. A pragmatic perspective explores “what works” using approaches which give priority to the importance of the research problem and question, valuing both objective and subjective knowledge (139).

This is an original piece of research undertaken for my PhD under the supervision of Professor John Browne. I performed the two systematic reviews which involved developing the search criteria, managing search retrievals, screening articles for possible inclusion, critically appraising relevant articles, selecting the most appropriate method of synthesis and performing the review. Professor Browne provided guidance and support throughout this process. Dr. Joanne Greenhalgh offered analytical support for the qualitative systematic review.

Professor Browne conceptualised the rationale for the trial. I formulated the proposal which included deciding the study design, identifying research participants and confirming measurement instruments. I sought copyright approval for the Oxford Hip Score, the EQ-5D and requested permission to use the Hip Osteoarthritis and Outcome Score. I was responsible for applying and receiving approval from 13 ethics committees to conduct the research. I organised and managed the data collection for the pre-feedback and post-feedback stage of the study. This involved training data collectors and monitoring the pre-operative data collection in each hospital. It also involved managing the post-operative data collection centrally from University College Cork (UCC). I coded all the questionnaires and was responsible for cleaning the data. I liaised with the Clinical Research Facility in UCC who performed the randomisation. I developed the feedback report and designed the educational session for the intervention. I performed the analysis for the feedback report and for the trial. I designed the semi-structured interviews, organised and facilitated the interviews, transcribed the recordings and performed the analysis. Dr. Carol Sinnott independently coded three randomly selected transcripts and helped to refine the framework prior to commencing the qualitative analysis. I was responsible for writing the thesis and for any publications emerging from this research.

This work is presented in the format of a collated thesis, comprising of four publications, each presented as a chapter which either have been or will be published by a relevant peer-reviewed academic journal. The papers are preceded by this introduction chapter and followed by a discussion chapter. Chapters 2, 3 and 5 have been published. Chapter 4 will be submitted for peer review after submission of the thesis.

1.3.4 Published papers to date

-Boyce, M.B. and J.P. Browne. Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients? A systematic review. *Qual Life Res*, 2013 (141).

-Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. *BMJ Qual Saf*. 2014 Feb 6 (142).

-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 (143).

2

Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients?

A systematic review

This article was published as:

Boyce, M.B. and J.P. Browne. Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients A systematic review. Qual Life Res, 2013 (141).

Chapter 2- Quantitative Systematic Review

2.1 Abstract

Purpose: To assess the impact of providing healthcare professionals with feedback on patient-reported outcome measures (PROMs).

Methods: This is a systematic review including controlled studies investigating the effectiveness of PROMs feedback, specifically examining the impact at a group-level and a patient-level.

Results: Only one study provided feedback at a group-level as a measure of professional performance, which found no intervention effect. At a patient-level, sixteen studies were identified and only one study found an overall significant difference in the PROM score. However, an additional six studies found a significant result favouring the intervention group for a particular subgroup or domain. The studies which demonstrated the greatest impact primarily used PROMs as a management tool in an outpatient setting on a specialised patient population. In contrast, there was weak evidence supporting the use of PROMs as a screening tool. The studies which found a positive effect had a lower quality score on average.

Conclusions: The effectiveness of PROMs feedback seems to be related to the function of the PROM. However, the effectiveness regarding the impact of PROMs feedback on patient outcomes is weak, and methodological issues with studies are frequent. The use of PROMs as a performance measure is not well investigated. Future research should focus on the appropriate application of PROMs by testing specific hypothesis related to cause and effect. Qualitative research is required to provide deeper understanding of the practical issues surrounding the implementation

of PROMs and the methodological issues associated with the effective use of the information.

2.2 Introduction

The use of patient-reported outcome measures (PROMs) in clinical effectiveness research is growing; however there is considerable uncertainty regarding their use in quality of care studies (144). PROMs quality of care studies involve feeding back information on patients' health, health-related quality of life, and other health-related constructs to professionals in an attempt to improve patient care. PROMs were not primarily developed for this purpose so little is known about how they perform in this context. Greenhalgh and Meadows strongly recommend that theories of change should be used to understand the mechanisms by which PROMs may stimulate changes in practice (145). PROMs have different functions depending on the level of aggregation of the data.

At a patient-level, PROMs are collected from patients and fed back to professionals usually during a consultation. In this situation, PROMs can act as: a screening tool for undiagnosed problems; a management tool to identify and prioritise issues; or a means to improve patient-physician communication (146). To understand the usefulness of PROMs feedback, it is necessary to examine whether it alters the decision making process. Greenhalgh *et al.* outlines that such feedback may initiate changes through: ordering further tests; referral to other professionals; changes in medicines or treatments; advice and education on better control or management of the problem. This information may promote better communication leading to a greater understanding of complex personal circumstances, joint decision-making, concordance through shared goals and greater patient satisfaction (147). It is

assumed that such improvements in the processes of care would ultimately impact positively on patients' health (112). Unfortunately, however, many of the existing PROMs are not adequately developed for individual-level comparisons (148).

At a group-level, patients' reports are aggregated and summary statistics are fed back to professionals. It is assumed that such feedback should stimulate more effective governance through performance monitoring, encourage clinical audit, and potentially influence resource allocation and policy decisions (85, 149). This method was first employed by Medicare (USA) and BUPA (UK) after recognising large variations in care (3, 73, 103). In 2009, the NHS introduced the use of PROMs as a national performance indicator in England. This made the collection of PROMs a mandatory requirement for four procedures (hip replacement, knee replacement, hernia repair and varicose vein surgery). Outcomes are compared at a hospital-level and reported publically to promote quality improvements (2).

The use of PROMs as a measure of the performance of healthcare professionals and organisations has a weak theoretical foundation. The rationale behind monitoring performance is to impose an inherent pressure to improve practice. This pressure is dictated by the approach adopted which can include monitoring alone, benchmarking, public-release of information and linking performance to incentives. Some believe that monitoring alone alters professional's performance through psychological pressures similar to the Hawthorne Effect (101). Benchmarking is based on the premise that professionals have an intrinsic competitive nature and peer comparison provides the necessary motivation to stimulate change (102). Public disclosure of outcomes data is thought to generate public accountability and market competition between professionals (66, 150). The use of incentives stimulates action by linking performance to rewards such as payment (151). PROMs feedback puts the

onus on healthcare providers to investigate the source of any variations observed. Therefore, there is an implicit assumption that performance monitoring will incentivise local data collection, in turn identifying areas for improvement and leading to better patient outcomes.

What is the current evidence?

There is some evidence that audit and feedback of clinical information are effective in improving professional practice. The effects are generally small to moderate but can be clinically meaningful (152). The public-release of performance data can stimulate quality improvements at the hospital-level (153). However, the effect of public disclosure on health outcomes is unclear; the evidence is scant, primarily focusing on mortality and cardiac care (153). In relation to patient-reported data, it is evident that patient experience measures can be effective in promoting quality improvements (154), but the literature on the impact of PROMs feedback remains inconclusive. The effectiveness of PROMs has been investigated in seven reviews since 1999, five of which were classified as systematic (83, 109-112, 145, 155). The reviews differ in aims, comprehensiveness and quality; however, consistent conclusions emerged across all reviews in relation to the lack of impact on outcomes (Table 3). The previous reviews focused on the value of feeding back patient-level PROMs data to healthcare professionals.

The current review

This review is the first to investigate the usefulness of PROMs feedback at both the patient- and group-level. The use of PROMs at the group-level is a major health policy initiative in England, but the evidence of their effectiveness in this context has never been synthesised. There is a large cost associated with collecting PROMs; therefore, evidence of the effect on patient outcomes is necessary to justify their use.

This review specifically investigates the effect of feedback on patient-reported health outcomes. Previous reviews have had a very broad focus, investigating the impact of feedback on a range of process and outcome measures. By measuring the impact of PROMs feedback with other measures, there is a chance that the effect of the intervention could be missed. Therefore, this review investigates the impact of PROMs feedback on the score of that particular PROM.

Furthermore, this review included studies which provided feedback to all healthcare professionals as clinicians no longer have sole responsibility over patient care, particularly since the introduction of multi-disciplinary teams (156-158). This review searched for all controlled designs to capture the full extent of the evidence and searched the literature up to 2012.

Table 3: Previous reviews on the effectiveness of PROMs feedback to healthcare professionals (n=7)

Review	Type of review (no. of studies)	Population	Intervention	Outcomes	Design	Comprehensiveness	Quality appraisal	Results	Data pooling	Issues and suggestions
Greenhalgh, 1999 (145)	Literature review (n=13) (7 included in this review)	Clinicians and patients	Impact of PROMs feedback to professionals	Process and outcome indicators: feasibility, acceptability, utility, management, satisfaction and health status	RCTs	Database searching (Medline, Cinahl, PsycLIT)	No	Information valued by clinicians, increased the detection of psychological and functional problems. Little impact on management or patient outcomes	No	Impact affected by implementation, population and setting, and outcome criteria Provide specific management guidelines, use disease-specific PROMs, link provision of feedback to patient visits
Espallargues, 2000 (109)	Systematic (n=21) (9 included in this review)	Individual or groups of physicians and patients	Impact of PROMs feedback to professionals	Process and outcome indicators: utilisation, diagnosis, treatment, health status and satisfaction	RCTs	Database searching (Medline), reference searching and contact with authors	Modified version of criteria proposed by Guyatt et. al and Sackett et. al	Impact on process of care but not patient functional or health status	Diagnosis, notation, or recognition of mental health issues and any change in prescribed medications or treatments	Interpretation of measures identified as an issue Successful studies used specific questionnaires. New or specific groups of vulnerable patients with active disease may benefit most

Review	Type of review (no. of studies)	Population	Intervention	Outcomes	Design	Comprehensiveness	Quality appraisal	Results	Data pooling	Issues and suggestions
Gilbody, 2001 (110)	Systematic (n= 9) (2 included in this review)	Clinicians and patients with anxiety and depression in a non-psychiatric settings	Impact of PROMs feedback to professionals	Process and outcome indicators: detection, initiation of treatment or referral, outcome of disorder, consulting behaviour, service use, patient satisfaction, communication and cost	RCTs	Database searching (Medline, Embase, PsycLIT, Cinahl, Cochrane Controlled Trials Register), hand-searched key journals and reference lists	Jadad scale	Increased recognition but no impact on patient management or outcomes	Recognition of depression	Questioned the sensitivity and specificity of PROMs, the willingness and ability of clinicians to deal with emotional disorders, and the unit of randomisation (contamination) Requires simple feedback and user friendly administration
Gilbody, 2002 (111)	Systematic (n= 9) (6 included in this review)	Clinicians and patients with mental health issues in non-psychiatric and psychiatric settings	Impact of PROMs feedback to professionals	Process and outcome indicators: detection, initiation of treatment or referral, outcome of disorder and changes in HRQoL,	RCTs and quasi-randomised trials	Database searching (Medline, Embase, PsycLIT, Cinahl, Cochrane Controlled Trials Register), hand-	Jadad scale, criteria of Schulz et. al and Cochrane criteria	Little impact on recognition, outcomes or clinical decision making	No	Questioned the use of measures at an individual-level as instruments are not designed for this purpose

Review	Type of review (no. of studies)	Population	Intervention	Outcomes	Design	Comprehensiveness	Quality appraisal	Results	Data pooling	Issues and suggestions
				consulting behaviour, service use, hospital status, patient satisfaction, communication and cost		searched key journals and reference lists of studies				
Marshall, 2006 (155)	Structured (n= 38) (10 included in this review)	Healthcare providers and patients	Impact of PROMs feedback to professionals	Process and outcome indicators: communication, concordance, provider and patient behaviours, patient satisfaction, health status, and resource use	RCTs and non-RCTs	Database searching (Medline) Hand-searched reference lists of studies and prior reviews	No	Impact on process of care (diagnosis, management, communication) but inconsistent effect on patients' health status	No	Questioned the value of PROMs given their different roles Multidimensional and individualised measures may be more useful
Valderas, 2008	Systematic (n=34) (10	Individual physicians or groups	Impact of PROMs feedback to	Process and outcome indicators:	RCTs	Database searching (Medline and	Modified Jadad scale	Impact on 65% of studies measuring	No	Methodological concerns limit the strength of evidence.

Review	Type of review (no. of studies)	Population	Intervention	Outcomes	Design	Comprehensiveness	Quality appraisal	Results	Data pooling	Issues and suggestions
(83)	included in this review)	of physicians and patients	professionals	mortality, morbidity, HRQoL, clinician behaviour, clinician impressions, patient satisfaction, and cost		the Cochrane Library), hand-searched reference lists of studies and prior reviews, contact with authors and experts		process of care and 47% of studies measuring outcomes of care		Studies need to consider the unit of randomisation (contamination) and statistical methods. Clinical staff should implement intervention
Lockett, 2009 (112)	Review (n=6) (1 included in this review)	Healthcare professionals and patients in the area of oncology	Impact of PROMs feedback to professionals	Outcome indicators: satisfaction, health status and resource use.	RCTs	Database searching (Medline and PsycINFO), plus included relevant articles from previous reviews	No	Impact on communication but effect on patient outcomes was limited	No	Information was not routinely used by clinicians Adequate training required, adopt individualised measures and Computer Adapted Testing. Use a clustered design to control for contamination effects

2.3 Methods

Eligibility criteria

Studies which met the following criteria were included: language of publication was English; participants receiving PROMs feedback were individual or groups of healthcare professionals; comparison was feedback (with or without educational support) versus no feedback; patients received normal care except for feedback of PROMs; outcome was measured by a change in PROMs score which had to be the same as the measure used to generate feedback; and studies involved a controlled design. These criteria were adapted from previous reviews (83, 109).

Search strategy

Forty-three articles were identified from previous systematic reviews and included for full-text evaluation. To capture additional studies, a search strategy was developed involving four blocks including: PROMs (adapted from a previously developed search strategy (88)); audit and feedback; professional competence; study design including Randomised Controlled Trials (RCTs) (adapted from a previously developed search strategy (159)) and non-randomised studies. No time restriction was placed on the search (Appendix 1). A professional librarian assisted in formulating this strategy. A search was performed in PubMed and the Cochrane Library in February 2012. In addition, reference lists of articles selected for full text review were screened, and a citation search of previous reviews was performed.

Full-text eligibility

An initial screening was performed using the participants, intervention, comparison and outcome (PICO) parameters. MB screened and reviewed all articles. This

involved undertaking a title and abstract search of articles identified. If there was a possibility that an article would fulfil the eligibility criteria, it was selected for full article evaluation.

Critical appraisal and data extraction

A modified version of the Jadad scale was used to appraise study quality as the original scale does not capture important dimensions of quality in this context and includes an item on blinding which is universally absent (160). Studies were scored on a 0-6 scale, with higher scores indicating better quality. The modified version assessed randomisation (up to 2 points), description of withdrawals/dropouts (1 point) and whether these were equal between groups (1 point), and the appropriate implementation of a cluster design to prevent contamination (up to 2 points). Fayer identified that contamination is an issue for PROMs intervention studies which randomise at the level of the patient. If healthcare professionals receive feedback for some patients, it is reasonable to assume that this will heighten their awareness regarding this issue for all patients diluting the intervention effect (161).

All articles that met the inclusion criteria underwent data extraction for information about study design, setting of study, participants, unit of randomisation (if applicable), intervention details, PROMs instruments, administration of PROMs, primary outcome measure and results. A second reviewer (JB) assessed the quality of data extraction and validity of findings in a randomly selected subset of five studies and was satisfied with the quality of the original data extraction exercise.

Analysis

A narrative synthesis of results was performed. A meta-analysis was proposed but insufficient data on the precision of results were provided. Results were reported in

different ways by studies including comparisons of the magnitude of change scores in intervention and control groups, and comparison of post-feedback scores alone. Effect sizes for individual studies were reported, where possible, using Cohen's *d* statistic. This divides the mean difference between groups by the pooled standard deviation. Effect sizes are categorized as: small (0.2), medium (0.5) and large (0.8) (162).

A study was considered 'positive' for PROMs feedback when there was an overall significant difference in PROMs scores between groups ($p < 0.05$). However, positive findings among domains and subgroups between arms were also reported, as certain health dimensions or patient groups may benefit more from the intervention.

2.4 Results

A total of 3,324 potentially relevant publications were identified. Many of the studies screened were descriptive examining the validity or feasibility of PROMs. Seventy-six full-text articles were reviewed: 56 of these were excluded either because the language was not English ($n=1$), the intervention was not PROMs feedback ($n=16$), the outcome was not a PROMs score ($n=20$), the intervention and outcome measure were not the same ($n=15$), or patients were treated differently across groups ($n=4$). Twenty studies met the inclusion criteria. Three studies with a non-randomised design which focused on patient-level feedback were excluded from the analysis due to the large number of randomised studies available for review (Figure 5 and Table 4).

This is an original systematic review. Of the 43 studies selected for full-text evaluation from previous systematic reviews, only 11 met the inclusion criteria for this review (163-173). Four studies were excluded as the intervention was not

PROMs feedback, 18 were excluded because the outcome was not a PROMs score, six were excluded because the intervention and outcome were not the same PROM, and four were excluded because patients were treated differently across groups. Six new studies which were not included in any of the previous systematic reviews are covered in this review. Two of these studies were captured because our review covers all healthcare professionals (174-175), one because our review did not exclude studies with non-standardised instruments (modified measures) (176), two because our review is more up to date (177-178) and one because we have included studies where PROMs feedback was provided at the group-level (179).

Figure 5: Flow chart of studies in quantitative review

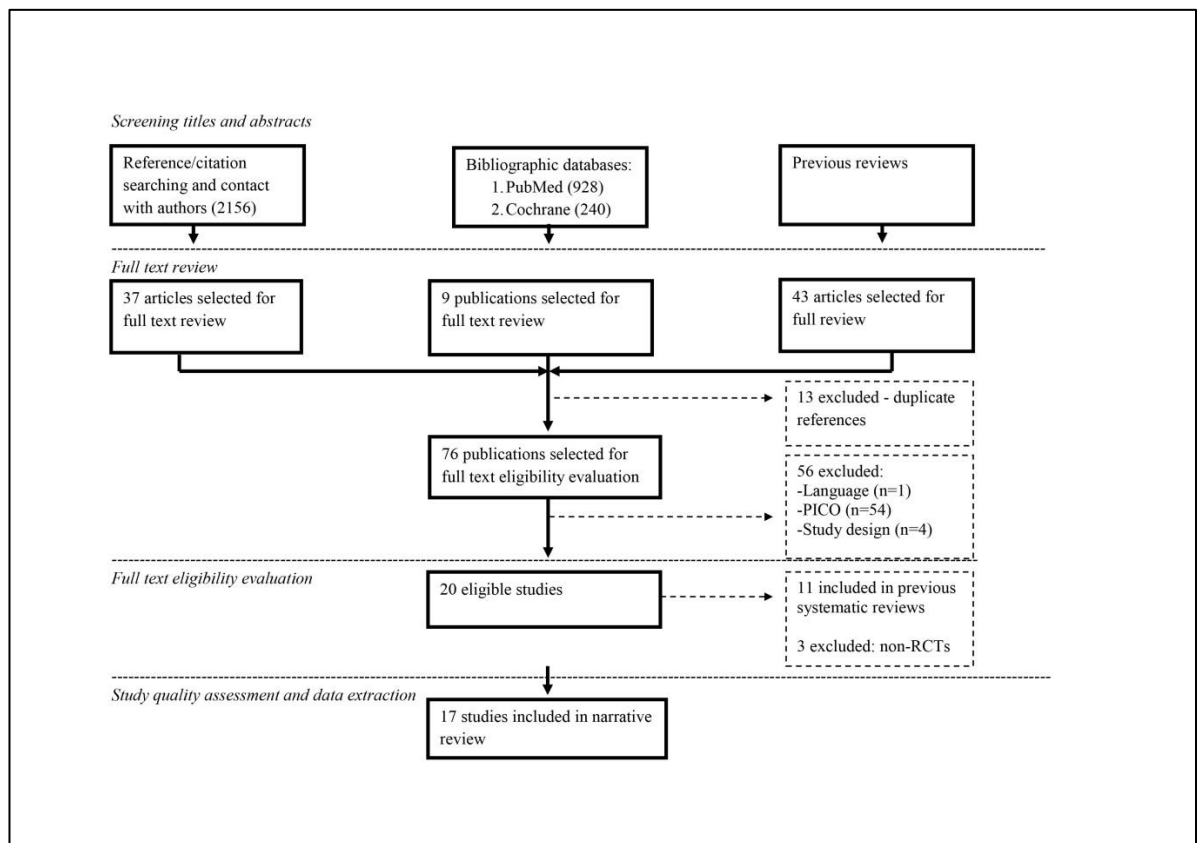


Table 4: Studies investigating the impact of PROMs feedback to healthcare professionals (n=17)

Ref	Design and allocation	Level of feedback and function	Patients (n)	Professionals (n)	Healthcare issue, setting, country	Sample size calculation	Intervention (frequency/format)	Control	Outcome	Quality score	Results and effect size	Internal validity
Calkins, 1994 (163)	Cluster RCT (2 groups) Randomised by practice teams (n=4)	Patient-level as a management tool	Existing patients, disabled adults (n=497)	Internists (n=8) & residents (n=52)	Functional disability, Outpatient care, USA	Not reported	Multiple feedback (x4) on FSQ (summary of questionnaire) and two-hour educational seminar	No feedback	FSQ (Generic)	3	Difference in number of bed days favouring intervention group	Selection and attrition bias
Dowrick, 1995 (164)	RCT (2 groups plus detected control group) Randomised by patient	Patient-level as a screening tool	Existing patients, adults (n=179)	GPs (n=9)	Depression, Primary care (n=2 practices), UK	To detect a difference of 15% on depression score	Single feedback on BDI (patient's details, depression score and diagnostic interpretation)	No feedback Median change (CI) = -3 (-5.5 to -0.5)	BDI (Dimension specific)	2	Both groups improved over the course of the study	Selection, attrition and measurement bias
German, 1987 (165)	RCT (5 groups: 2 intervention and 3 control) Randomised	Patient-level as a screening tool	Existing patients, adults (n=809)	GPs and residents (n= 45)	Depression, Primary care, USA	Not reported	Single feedback on GHQ (total score, subscale scores, positively answered)	No feedback	GHQ (Dimension specific)	1	No significant differences between groups	Attrition and measurement bias

Ref	Design and allocation	Level of feedback and function	Patients (n)	Professionals (n)	Healthcare issue, setting, country	Sample size calculation	Intervention (frequency/format)	Control	Outcome	Quality score	Results and effect size	Internal validity
	by patient						items and explanation of GHQ)					
Gutteling, 2008 (178)	Cluster RCT (2 groups) Randomised by physician	Patient-level as a management tool	Patients volunteered to participate, adults (n=162)	Physicians (n=11)	Liver disease, Outpatient care, The Netherlands	To detect a minimum effect size of 0.5 (did not specify which measure this calculation was based on)	Multiple feedback on LDSI (graphical output including previous measurements) and educational session on interpretation of output	No feedback	Severity items of LDSI (Disease specific) SF-12 (Generic)	3	Subgroup effect (age and gender) favouring intervention group *Effect size = 0.15	Selection and attrition bias
Hawkins, 2004 (175)	RCT (3 groups) Randomised by patient	Patient-level as a management tool	All patients in therapy, adults (n=201)	Psychologists (n=3) and licensed social workers (n=2)	Mental health, Outpatient care based psychiatric clinic, USA	Not reported	Multiple feedback on OQ-45 (graph conveying progress and recommendations)	No feedback Mean change (SD)= 14.39 (16.61)	OQ-45 (45 item: Dimension specific)	3	Overall positive effect favouring intervention group *Effect size = 0.28	Attrition and measurement bias

Ref	Design and allocation	Level of feedback and function	Patients (n)	Professionals (n)	Healthcare issue, setting, country	Sample size calculation	Intervention (frequency/format)	Control	Outcome	Quality score	Results and effect size	Internal validity
Kazis, 1990 (166)	RCT (3 groups: 2 studies reported) Randomised by patient	Patient-level as a management tool	All patients, adults. Study 1 n= 710 Study 2 n=1210	Specialists (n=?)	Rheumatoid arthritis, arthritis centres, USA (n=2) Study 1 (n=12 clinics) Study 2 (n=15 clinics)	Based on small to moderate effects	Multiple feedback (x3) on AIMS and MHAQ (single page)	No feedback	Study 1: AIMS (Disease specific) Study 2: MHAQ (Dimension specific)	2	No significant differences between groups	Selection and measurement bias
Lambert, 2001 (174)	RCT (4 groups) Randomised by patient	Patient-level as a management tool	Consecutive patients, adult (n= 609)	Counseling centre staff: PhD level psychologists (n=16), doctoral students including interns (n=15)	Mental health, Counseling centre, USA	Not reported	Multiple feedback on OQ-45 (graph conveying progress and recommendations)	No feedback	OQ-45 (45 item: Dimension specific)	1	Subgroup effect (patients not on track) favouring intervention group	Attrition and measurement bias

Ref	Design and allocation	Level of feedback and function	Patients (n)	Professionals (n)	Healthcare issue, setting, country	Sample size calculation	Intervention (frequency/format)	Control	Outcome	Quality score	Results and effect size	Internal validity
Mathias, 1994 (167)	Cluster RCT (2 groups) Randomised by call group	Patient-level as a screening and management tool	Patients with unrecognised and untreated anxiety, adults (n=573)	Physicians (n=75)	Anxiety, Primary care (n=23 practices), USA	Not reported	Multiple feedback on the SCL-90-R and SF-36 (displaying patients' profiles) and educational session on interpretation of instrument and management of anxiety	No feedback	SCL-90-R (Dimension specific) SF-36 (generic)	2	Both groups improved over the course of the study	Selection and attrition bias
McCoy, 1988 (176)	Cluster RCT (2 groups) Randomised by physician	Patient-level as a screening tool	New patients, adults (n=608)	Resident physicians (n=74)	Functional disability, Ambulatory clinic, USA	Not reported	Single feedback on FSQ (graphical including warning zones and analysis of specific problems linking to FSQ)	No feedback	Expanded FSQ (64 items: generic) Chronic disease inventory (Disease specific)	2	No significant differences between groups	Selection and attrition bias

Ref	Design and allocation	Level of feedback and function	Patients (n)	Professionals (n)	Healthcare issue, setting, country	Sample size calculation	Intervention (frequency/format)	Control	Outcome	Quality score	Results and effect size	Internal validity
Puschner, 2009 (177)	Cluster RCT (2 groups) Randomised by clinician	Patient-level as a management tool	All inpatients, adults (n=294)	Clinicians (n=45)	Mental health, Inpatient (n=10 wards), Germany	To detect a medium effect (0.5 standard deviation units on OQ-45). Accounted for attrition and inflated for cluster design	Multiple feedback on EB-45 (single page with graphs showing progress and recommendations) and educational session including quality circles to discuss feedback	No feedback	EB-45 (45 items: Dimension specific) <i>German version of the OQ-45.</i>	6	No significant differences between groups *Effect size: = -0.14	Attrition bias
Rubenstein, 1989 (168)	Cluster RCT (2 groups) Randomised by practice	Patient-level as a management tool	Existing patient, adults (n=510)	GPs and internist (n=76)	Functional disability, Primary care, USA	Not reported	Multiple feedback (x4) on Beth Israel UCLA FSQ (one page displaying scale scores, warning levels and narrative summary) and two hour	No feedback	Beth Israel UCLA FSQ (34 items: Generic)	3	No significant differences between groups	Selection and attrition bias

Ref	Design and allocation	Level of feedback and function	Patients (n)	Professionals (n)	Healthcare issue, setting, country	Sample size calculation	Intervention (frequency/format)	Control	Outcome	Quality score	Results and effect size	Internal validity
							multimedia educational session					
Rubenstein, 1995 (169)	RCT (2 groups) Randomised by clinic module	Patient-level as a screening tool	All new patients, adults (n=557)	General internists residents (n=73)	Functional disability, Outpatient clinic, USA	Not reported	Single feedback on FSQ (graphs and narrative summary) and half hour educational session, plus booster session at three months	No feedback	FSQ (generic)	2	Significant difference in mental health scores favouring intervention group	Selection and attrition bias
Trowbridge, 1997 (173)	RCT (2 groups) Randomised by patient	Patient-level as a management tool	Existing patients, adults (n=510)	Oncologists (n=13)	Oncology, Outpatient clinic, USA (n=23)	Not reported	Single feedback on pain level (raw data)	No feedback	Pain level	1	Significant decrease in usual aches and pains favouring intervention group	Selection, attrition and measurement bias

Ref	Design and allocation	Level of feedback and function	Patients (n)	Professionals (n)	Healthcare issue, setting, country	Sample size calculation	Intervention (frequency/format)	Control	Outcome	Quality score	Results and effect size	Internal validity
Wasson, 1999 (172)	Cluster RCT (2 groups) Randomised by practice (n=22)	Patient-level as a management tool	Existing elderly patients (>70) (n=1651)	Family practitioners (n=27) and internists (n=18)	General health, Primary care, USA	Not reported	Single feedback on activities of daily living and SF-36 (summarized on a flow sheet)	No feedback	Activities of daily living SF-36 (Generic)	4	Significant difference for instrumental activities of daily living favouring intervention group	Selection and attrition bias
White, 1995 (170)	Cluster RCT (2 groups) Randomised by practice (n=23)	Patient-level as a management tool	Existing patients with asthma (n=818)	Practitioners and nurses (n=?)	Asthma, Primary care, UK	Not reported	Multiple feedback (x4) on symptom frequency (summary and asthma index to develop thresholds) and educational session for professionals	No feedback	Frequency of symptoms (27 items)	3	Both groups improved over the course of the study	Attrition bias
Whooley, 2000	Cluster RCT (2 groups) Randomised	Patient-level as a screening	Existing patients (n=331)	GPs and internists (n=?)	Mental health, Primary care,	1.4 point difference in mean	Single feedback on GDS (scores	No feedback Mean	GDS (15 items-Disease	4	Both groups improved	Attrition bias

Ref	Design and allocation	Level of feedback and function	Patients (n)	Professionals (n)	Healthcare issue, setting, country	Sample size calculation	Intervention (frequency/format)	Control	Outcome	Quality score	Results and effect size	Internal validity
(171)	by clinic (n=13)	tool			USA	GDS scores (20% difference at two year follow up). Accounted for attrition.	and severity) and one hour session on management of depression to professionals and patient sessions on coping with depression	change (CI)= -2.1(1.5 to -4.2)	specific)		over the course of the study	
Group-level feedback												
Weingarten, 2000 (179)	Cluster RCT (2 groups) Randomised by physician	Group-level as performance indicator	Existing elderly patients (n=1810)	Internist, family practitioners, subspecialists (n=48)	Functional disability, Primary care, USA	Not reported	Single feedback on Darmount Primary Care Cooperative Information Project chart (aggregated peer comparison feedback)	No feedback	Darmount Primary Care Cooperative Information Project chart (Generic)	2	Functional status reduced for both groups	Selection and attrition bias

FSQ Functional Status Questionnaire; BDI Beck Depression Inventory; GHQ General Health Questionnaire; LDSI Liver Disease Symptom Index; SF-12 Medical Outcome Study Short Form 12; OQ-45 Outcomes Questionnaire 45; AIMS Arthritis Impact Measurement Scale; SCL-90-R Symptoms Check List-90-Revised; EB-45 German version of the Outcome Questionnaire 45; UCLA FSQ University of California, Los Angeles Functional Status Questionnaire; GDS Geriatric Depression Scale.

*Effect size formula= mean (control) - mean (Intervention)/ pooled standard deviation

Study setting

Thirteen of the 17 included studies were carried out in the USA, two were conducted in the UK, one in The Netherlands and one in Germany. Eight of the studies took place in an outpatient clinic, eight were set in primary care, and only one was based in an inpatient setting.

Study population

Most studies assessed the impact of feedback to physicians only (n=14) including general practitioners, internists, specialists and residents. Other healthcare professionals involved were as follows: psychologists, psychology doctoral students, nurses and licensed social workers (n=3). The healthcare issues covered were mental health (n=7), functional status (n=5), rheumatoid arthritis (n=1), oncology (n=1), general health (n=1), asthma (n=1) and liver disease (n=1). All studies focused on an adult population.

Study Design

All of the included studies were RCTs. Eleven of these used a cluster randomised design of which four were randomised by physician and seven randomised by team, clinic or practice.

The included studies were of generally poor quality (160). The mean quality score was 2.6 with individual study scores ranging from 1 to 6. Six studies did not attempt to deal with contamination between the intervention and control groups in the study design. Twelve studies did not provide a formal sample size calculation which increases the risk that real effects have been missed by some studies.

The method used to statistically compare control and intervention groups in studies included within-group change scores (n=7) (166-167, 169, 171-172, 175, 179) and

between-group post-feedback scores (n=8) (163, 165, 168, 170, 173-174, 177-178).

The method used in two studies was not clear (164, 176).

Intervention

The function of the PROMs differed across studies. Only one study fed back the information at a group-level as a measure of professional performance. At a patient-level, ten studies used PROMs as a tool to manage patient care by monitoring disease progression and assessing the effect of treatment, five used PROMs to screen for a specific issue (n=5), and one used the measure as both a screening and management tool (n=1).

The instruments used to provide feedback included dimension-specific (n=7), generic (n=5), disease-specific (n=2), or a combination of specific and generic measures (n=3). Only one study provided feedback at a group-level, the remaining 16 studies provided feedback at the individual patient-level. Of the latter, only ten indicated that feedback corresponded with a patient's visit, three provided no evidence of patient-physician correspondence during the study period (163, 170, 172), and three reported insufficient information on the intervention (166-168). Eight studies provided feedback alone, and eight provided feedback and an educational session. The educational session was provided primarily to professionals (n=7) but also to professionals and patients (n=1).

Presentation of the feedback differed considerably in content and format. Most studies provided scores with some level of explanation (n=16) and only one provided raw scores (173). A graphical representation of results was provided in seven studies. Studies benchmarked scores using the patients' previous scores (n=6), reference scores (n=2), summary scores (n=1), and average patient scores (assessed at the same time) across peers (n=1). In addition, three studies provided recommendations

and guidelines. Studies varied in frequency of feedback providing feedback on a single occasion (n=8) and multiple occasions (n=9).

Administration of PROMs occurred primarily in the clinic (n=9). Patients self-completed questionnaires in the majority of studies (n=15) using the traditional pencil-and-paper method (n=15) and computerised administration (n=2).

Impact of the intervention

Impact by function

The study which provided feedback at a group-level found no statistical difference between intervention and control groups (179). Only one of the studies which provided patient-level feedback found an overall significant effect of feedback. In this study, the PROM was used as a management tool to monitor disease progression (175). An additional six studies found a significant result favoring the intervention for a subgroup of patients (n=2) or for particular domains (n=4). Five of these used the PROM as a management tool and only one used the PROM as a screening tool (163, 169, 172-174, 178). In addition, four studies reported an improvement in both arms of the trial (164, 167, 170-171).

An effect size could only be calculated for one of the studies with a 'positive' effect. This found a small effect size (0.28) in favour of PROMs feedback (175).

Impact by setting

The study which found an overall positive effect of PROMs feedback was based in an outpatient setting. Five of the studies which found a positive effect for a particular subgroup or domain were also based in an outpatient setting. Only one of the eight studies based in a primary care setting found a positive effect. However, the four studies which found an improvement in both intervention and control groups were

based in primary care settings. The study based in an inpatient setting did not find any significant differences between groups.

Impact by healthcare issue

Of the studies which found at least one positive effect (n=7), six focused on patients with specific health issues and one on a more general population. The studies which found an improvement in both groups focused on screening for mental health issues in general primary care patients (n=3) and managing symptoms for patients with asthma (n=1).

Impact by professional

Of the seven studies with a positive effect, professionals included specialists (n=2), family practitioners (n=1), interns and residents (n=2) and the remainder included counsellors (n=2). For those that found no effect (n=6), the professionals included general practitioners (n=3), resident physicians (n=1) and specialists (n=2). The studies which found an improvement in both groups included general practitioners (n=4).

Impact by PROMs

The study which found an overall effect used a dimension-specific PROM. The studies which found an effect in subgroups and domains used dimension-specific (n=2), disease-specific (n=1) or generic PROMs (n=3).

Impact and quality of studies

The study which found an overall difference between groups had a quality score of 3. The six studies which found positive effects in subgroups or domains had an average quality score of 2.3, and the four studies which found an improvement in both groups had an average quality score of 2.7. The six studies which found no

effect between groups had an average score of 2.7. The study with the highest quality score (6) found no effect. No relationship was evident between the type of analysis and effect.

2.5 Discussion

This review investigated the impact of providing PROMs feedback to healthcare professionals on patient-reported outcomes. The one study which provided PROMs feedback at a group-level did not find an effect and in fact, found that health deteriorated for all participants. This study focused on the functional status of an elderly population over the course of the four years so the likelihood of finding an effect may have been outweighed by the level of health decline (179). Of the 16 studies which examined the value of feedback at the patient-level, only one found a statistically significant difference between the intervention and control groups. An additional six of these studies found positive results favouring the intervention group for certain domains or subgroups. The quality of the studies reviewed was generally poor. The studies which demonstrated some effect of feedback were of slightly lower quality than those which demonstrated no effect. This raises the possibility that any positive benefits attributed to the use of PROMs feedback may be due to study biases.

There is tentative evidence that the effectiveness of PROMs feedback is related to certain study features. Studies which used PROMs as a management tool (primarily based in an outpatient setting on a specialised patient population) demonstrated the greatest impact of feedback. There was less data supporting the use of PROMs to screen for otherwise unsuspected conditions as only one study that evaluated the use of PROMs in this context found a statistically significant benefit. It is possible that

the impact on specific domains and subgroups may be due to multiple testing. However, it is also possible that feedback may be more effective for specific populations or healthcare issues. Espallargues *et al.* referred to this in a previous review suggesting that new or specific groups of vulnerable patients with active disease may benefit most from feedback (109). Reasons why PROMs as a management tool on specific patients may be more effective may be: the actionability of the feedback by focusing on a specific problem the solutions for improvement may be clearer; the variability of outcomes as specialised patients have more severe symptoms, so the room for improvement and the potential to have an impact on patients' health may be greater.

Studies based on screening primary care patients tended to be less effective.

Whooley *et al.* questions whether participants had a clinically significant issue in the first instance as screening tools can over-exaggerate incidence (171). In addition, an effect would not be found if the study population had stable health conditions or if a high standard of care was already being provided. Furthermore, the benefits of screening general primary care patients may not be evident within the scope of the study period, compared to the short-term impact of managing specific patients with more severe and obvious symptoms. The same is true for the use of PROMs at a group-level where the effect of feedback may take years to substantially alter professional practice and ultimately filter through to patient outcomes. Lastly, the benefit of feedback will only be realised if the implementation of the intervention is effectively facilitated, including adequate buy-in from senior staff, data collection and technical support, and active use of the information. Puschner *et al.* recognised the limitation of their study by failing to incorporate strategies which encouraged clinicians to actively use the feedback (177).

Four studies found an improvement in outcomes for both groups in the trial (164, 167, 170-171). This may be because the completion of the questionnaire had a therapeutic effect on patients or made patients aware of the problem empowering them to seek medical advice independently (166, 177).

It is striking that although the systematic reviews in this field have different aims, they have all concluded that the impact of PROMs feedback on patient outcomes is weak and methodological issues are frequent. The proportion of studies in previous reviews which found a significant difference on PROMs scores ranged from 0 to 60% (110, 112). This review falls within a similar range as 41% of studies found a positive effect.

This review has some limitations. The review is limited to the published data which tends to be of low quality. A meta-analysis was not possible as only three studies reported sufficient information. The populations, settings and interventions of these studies were too heterogeneous to conclude that the real variability was due to the intervention and not the variability within the study designs. A meta-regression between studies which provided feedback at a patient-level and group-level was also not possible, as only one study was identified in the latter category. The search focused on two bibliographic databases. However, an extensive search of references and citations was also performed yielding many additional studies. While this review only focused on English-language articles, the potential for bias was low as only one non-English article was identified (180). A problem with many of the feedback mechanisms observed in the literature is the over-reliance on the use of pencil-and-paper methods which limit the richness of feedback that can be delivered in a short space of time. Future studies should evaluate the effectiveness of specialised decision support software, which delivers more timely and sophisticated feedback to

professionals. Finally, the review has a narrow focus investigating the impact of PROMs feedback on the PROM itself. This has two consequences. First, benefits such as provider-patient communication are not captured and second, we focus only on a subset of the literature. However, we feel that our review provides an original contribution because we focus on the literature that specifically addresses the value of recent policy initiatives such as the NHS PROMs Programme.

In conclusion, the use of PROMs as a quality improvement tool is a highly versatile and complex intervention. This review has identified possible effective features and also an obvious gap in the literature on the value of group-level PROMs data as a performance assessment tool. This latter finding highlights a consistent failure by health systems when implementing national policies; resources are invested in policies which are not based on evidence of effectiveness and are not underpinned by a theoretical basis on the mechanisms by which change may occur. As a consequence, much ambiguity surrounds the primary objective of such policies resulting in little evaluation once enforced. The use of PROMs as a performance measure is currently receiving much interest from policy-makers internationally. Currently, the PROFILE trial based in Ireland is evaluating the effectiveness of PROMs feedback as a measure of surgeons' performance and will add to the evidence base. Other trials in this field should be prioritised by research funding agencies.

Although, there is a body of qualitative literature focusing on the value of PROMs feedback to healthcare professionals, there has been no previous attempt to synthesise this evidence. Qualitative evidence can help us understand barriers and facilitators to change, and problems with the impact of complex interventions. There

is a need to examine this evidence to provide a deeper insight into the use of PROMs as quality improvement tools.

3

**The views of professionals who are using Patient-Reported
Outcome Measures as quality improvement tools:
A systematic review of qualitative research**

This article was published as:

Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. BMJ Qual Saf. 2014 Feb 6.

Chapter 3- Qualitative Systematic Review

3.1 Abstract

Objectives: To synthesise qualitative studies that investigated the experiences of healthcare professionals with using information from patient-reported outcome measures (PROMs) to improve the quality of care.

Design: A qualitative systematic review was conducted by searching PubMed, PsycINFO and CINAHL with no time restrictions. Hand searching was also performed. Eligible studies were evaluated using the Critical Appraisal Skills Programme toolkit for qualitative studies. A thematic synthesis identified common themes across studies. Study characteristics were examined to explain differences in findings.

Setting: All healthcare settings.

Participants: Healthcare professionals.

Outcomes: Professionals' views of PROMs after receiving PROMs feedback about individual patients or groups of patients.

Results: Sixteen studies met the inclusion criteria. Barriers and facilitators to the use of PROMs emerged within four main themes: collecting and incorporating the data (practical), valuing the data (attitudinal), making sense of the data (methodological) and using the data to make changes to patient care (impact).

Conclusion: Professionals value PROMs when they are useful for the clinical decision-making process. Practical barriers to the routine use of PROMs are prominent when the correct infrastructure is not in place before commencing data collection and when their use is disruptive to normal work routines. Technology can

play a greater role in processing the information in the most efficient manner.

Improvements to the interpretability of PROMs should increase their use. Attitudes to the use of PROMs may be improved by engaging professionals in the planning stage of the intervention and by ensuring a high level of transparency around the rationale for data collection.

3.2 Introduction

Patient-reported outcome measures (PROMs) are questionnaires that assess patients' health, health-related quality of life, and other health-related constructs (1). They have traditionally been used to describe the burden of disease and to establish the comparative effectiveness of different treatments (3). There is increasing interest in the use of PROMs to improve health services. Many policy makers and researchers believe that PROMs provide an essential perspective on the quality of health services (2-3, 73) and it has been suggested that they have the potential to transform how healthcare is organised and delivered (72). PROMs have been used to compare and reward the performance of healthcare providers in England (3), America (103-104), Australia (105-107) and Sweden (104), and their potential to improve quality has also been recognised in Canada (73) and the Netherlands (108).

The mechanisms through which PROMs feedback to healthcare professionals might improve the quality of healthcare depends on the type of feedback provided.

PROMs may be used to provide professionals with information about their performance against their peers (1, 3). It is posited that PROMs should act to improve the quality of healthcare in the same way as any other benchmarking tool (2-3). Peer benchmarking is thought to stimulate an intrinsic desire in healthcare professionals to succeed relative to their peers (102). In addition, it is hypothesised

that professionals and organisations are motivated to avoid any negative consequences of peer benchmarking. These consequences depend on the extent to which the benchmarking exercise is used to support broader quality improvement strategies such as clinical governance, payment by performance, clinical commissioning and patient choice (3, 181). For example, PROMs are used alongside other indicators to measure the performance of English National Health Service (NHS) providers and drive up quality throughout the NHS “by encouraging a change in culture and behaviour focused on health outcomes not process” (182). PROMs are also used in England to guide the award of ‘bonus’ payments to NHS Trusts (183), to inform the decisions of commissioning bodies about which NHS Trusts to contract with (184) and to facilitate patients when choosing a provider for certain elective surgical procedures (185). Finally, it is hypothesised that although the benchmarking of outcomes does not provide a direct insight into the causes of inter-professional performance variation, it can stimulate audit and research activities that might lead to the discovery of these causes. For example, professionals who are discovered to have poor performance might learn from the practices of those with the best performance (186).

Patient-level PROMs feedback can also be provided to professionals. This is hypothesised to facilitate personalised care management by highlighting the concerns and needs of individual patients in a structured format (81). The information can be used to highlight previously unrecognised health problems (169), assess the effectiveness of different treatment plans (174), monitor disease progression (178), stimulate better communication (187) and promote shared decision making (175, 188). Specific quality improvements that might arise from a consideration of PROMs feedback include ordering additional tests, referring the

patient to a new specialist, amending prescribed medicines or treatments, issuing personalised advice and education on symptom management, and altering the goals of treatment plans to better reflect patient concerns (82, 147).

The evidence supporting the effectiveness of PROMs in contributing to improvements in the quality of healthcare is heterogeneous, and it has been difficult to draw definitive conclusions about their impact on patient care (141). While there is some evidence that PROMs are effective in enhancing patient-clinician communication and helping to recognise new health issues, there is little evidence that PROMs feedback to healthcare professionals changes care management or improves patient outcomes (141, 189). This evidence should be considered alongside findings from the broader literature. First, the effects of audit and feedback interventions are generally small to moderate and we understand relatively little about the complex process dynamics associated with successful interventions (30). Second, the use of theory in studies of audit and feedback is rare which signals a need for more theoretically informed interventions (49).

Qualitative research with end users plays an important role in helping us understand why interventions are ineffective in practice and in the development of theoretical models to support successful implementation. Examining first-hand experiences may provide unique insights into the challenges associated with implementing and using PROMs in practice (146, 190). Synthesising this evidence may help explain the modest impact of PROMs on professionals' behaviour to date. Two previous reviews have reported the evidence about professionals' views on the use of outcome measures in general, not specifically focusing on PROMs (191-192). The first was a non-systematic review that provided an overview of the barriers to the routine use of outcome measures (191). The second was a systematic review which looked at the

barriers and facilitators to the use of outcome measures in routine practice (192).

This review was limited to the views of allied health professionals and excluded professions such as medicine and nursing. Given the unique methods and perspectives introduced by PROMs, and their broad use across different professional groups, there is a clear need for a systematic review of the qualitative literature that focuses exclusively on PROMs and includes all relevant healthcare professionals.

This review aimed to identify qualitative studies that have investigated the experiences of healthcare professionals with the use of PROMs as a means to improve the quality of healthcare and to synthesise findings about the barriers and facilitators to their use. The review also explores how the characteristics of different studies influenced the results observed.

3.3 Methods

Eligibility criteria

Studies that met the following criteria were included: language of publication was English; participants were healthcare professionals; examined professionals' views of PROMs after receiving PROMs feedback about individual patients or groups of patients; and used a qualitative design.

Information sources

A search without time restriction was performed in PubMed, PsychINFO and CINAHL in August 2013 (Appendix 2). Reference lists of included papers were screened for additional studies.

Search

A search strategy was developed comprising three blocks of terms relating to PROMs, qualitative research and professionals' opinions. Brettle *et al.* previously developed a comprehensive filter for PROMs, which was used as the first block for this search (193). The second block was based on a published search filter developed to capture qualitative evidence (194). The third block was developed by the authors to meet the aims of this specific review. It combined terms relating to 'professionals' and 'opinions', and used a proximity operator which identified any combination of these terms when they appeared within three words of each other.

Study selection

MB initially screened the titles and abstracts of articles retrieved by the search strategy. The full text of potentially relevant articles was evaluated if there was not enough information to make an informed decision about relevance to the systematic review from the abstract. Where there was continued uncertainty about whether such papers met the inclusion criteria, another reviewer (JB) was consulted for a second opinion and discrepancies were discussed to form a consensus.

Data collection process

All articles that met the inclusion criteria underwent data extraction for information about study aims, location and setting, study design, participants, recruitment, PROMs used, level of application, feedback strategy and study findings. A quality appraisal of included studies using an established toolkit was performed by MB, and reviewed by JB (195). The quality appraisal assessed the following criteria: appropriate design, appropriate recruitment strategy, appropriate data collection method, reflexivity, ethical research, appropriate analytic method, appropriate

discussion of findings and overall value. A sensitivity analysis was performed using matrices to compare the patterns of themes identified in studies of different quality.

Synthesis of results

Thematic synthesis was used to analyse the papers included in the review (196). It compares themes across studies, looks at study characteristics to help explain differences in findings and develops interpretations beyond original studies to generate analytical themes (196). The synthesis was performed by entering the entire results section from each study into QSR International's NVivo 10 software (197). The synthesis involved three stages: free line-by-line coding of findings from primary studies, categorising free codes to develop descriptive codes and developing analytical themes which explored the relevance of the descriptive codes in the context of the research question (196). Study characteristics and findings were cross-referenced on a matrix to explore whether thematic patterns were associated with certain studies. Meetings and correspondence between the co-authors throughout the analysis process helped to evolve the themes and challenge the interpretation of the data.

3.4 Results

Study selection

In total, 8,344 potentially relevant publications were identified by our search strategy and 7,930 were excluded on the basis of their titles. An abstract review of the remaining 414 articles was performed and 87 were chosen for full-text review. Seventy-one articles were excluded at the full-text stage leaving 16 relevant articles (Figure 6 and Table 5). These were an entirely different set of studies to those

included in the only previous systematic review of professional opinions about the routine use of outcome measures (192).

Figure 6: Flow chart of studies in the qualitative review

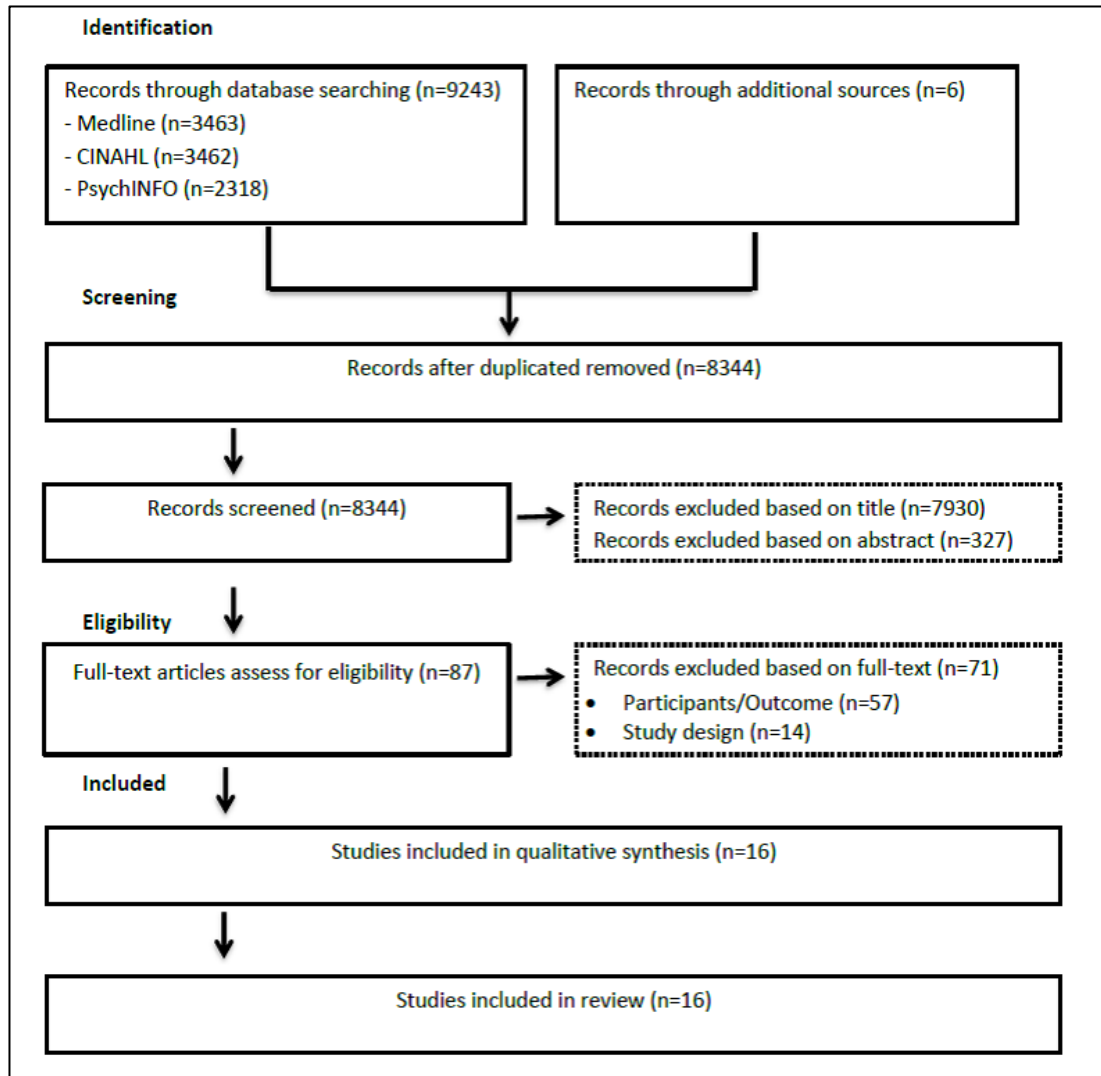


Table 5: Studies investigating the views of professionals (n=16)

Reference	Location, setting and focus	Study design	Participants	PROMs feedback	Study Aims
Bendtsen, 2003 (78)	Sweden, hospital setting, COPD	Focus groups (n=2)	Physicians (n=9)	Patients completed SF-36 on a touch screen computer and feedback was provided during the consultation	‘To examine the thoughts and attitudes among physicians concerning the value of an HRQoL measurement in addition to the traditional clinical and laboratory data used’
Callaly, 2006 (105)	Australia, public mental health service	Focus groups (n=13) and interviews (n=7)	Nurses (n= 64) Allied health professionals (n=12) Medical staff (n=7)	Patients completed BASIS-32 on a computer generating immediate feedback for professionals. Aggregated data reported publically	‘This paper explores the attitudes of mental health workers in one public health service towards the implementation and use of routine measurement’
Cranley, 2004 (198)	Canada, hospital setting, acute care	Informal semi-structured interviews	Nurses (n= 29)	Continuous assessment and feedback of information on functional status, symptoms, therapeutic self-care, falls and pressure ulcers	‘To provide initial insight from rational and phenomenological theoretical perspectives into how nurses integrate baseline and follow-up outcomes assessment into practice to inform their clinical decision-making’
Dorwick, 2009 (199)	UK, primary care, depression	Semi-structured interviews	GPs (n= 34)	Patients completed PHQ-9, HAS or BDI and feedback was provided immediately to GPs	‘To gain an understanding of doctors’ and patients’ views of the introduction of severity questionnaires for depression and their implementation in practice’
Dunckley, 2005 (200)	UK, nursing home and hospice, palliative care	Action research including interviews	Nurses (n=8) Doctor (n=1) Health care assistants (n=6)	Unclear details on feedback. POS collected from patients and clinicians	‘To further understand the barriers to outcome measure implementation and to identify and facilitate methods of over-coming these hurdles’

Reference	Location, setting and focus	Study design	Participants	PROMs feedback	Study Aims
Eischens, 1998 (201)	US, hospice setting, palliative care	Interviews	Nurses (n=8)	Patients completed McGill and HQLI, and feedback was provided immediately to nurses	‘The purpose of this study was to assess whether hospice nurses found QOL evaluations useful in designing and adjusting their patients care plans’
Hughes, 2003 (202)	UK, palliative care	Semi-structured interviews	Professionals (n=22)	Patients and staff completed POS, and feedback was provided to staff	‘The objective of this study was to elicit professional views and experiences of using outcome measures’
Hughes, 2004 (203)	UK, hospital, nursing home and primary care setting, palliative care	Semi-structured Interviews	Staff (n=13 of which 12 were nurses)	Patients and staff completed POS, and feedback was provided immediately to staff	‘The study aimed to: describe the implementation of a palliative care outcome measure in non-specialist palliative care setting and to understand the implementation of the setting’
Kettis-Lindblad, 2007 (204)	Sweden, hospital setting, oncology	Semi-structured interviews	Oncologists (n=6)	Patients completed SEIQoL-DW and disease-related SEIQoL on touch-screen computer, and feedback was provided during the consultation	‘This study explored patients’ and oncologists’ perceptions of using a computer-administered, individualised QOL instrument to support an oncologic consultation’

Reference	Location, setting and focus	Study design	Participants	PROMs feedback	Study Aims
Mason, 2008 (205)	UK, primary care, post-natal depression	Semi-structure interviews	Health visitors and nurses (n=19)	Patients completed EPDS and feedback was provided immediately to GPs	To address beliefs behind attitudes using a qualitative methodology to access the perceptions of healthcare professionals towards screening using the EPDS
Meehan, 2006 (106)	Australia, mental health setting	Focus groups (n=34)	Mental health staff (n=324)	Patients completed Mental Health Inventory on a computer generating patient level feedback or summary reports for comparisons (clinician reported measures also collected)	‘The aim of this study was to explore clinician reactions to (i) the introduction of routine outcome measures and (ii) the utility of outcomes data in clinical practice’
Mitchell, 2011 (206)	UK, primary care, depression	Focus groups (n=4)	Multi-disciplinary teams including GPs, nurses, doctors in training, mental health workers and managers (n=38)	Patients with new-onset depression completed PHQ-9 and feedback was provided immediately to professionals	‘To explore primary care practitioner perspectives on the clinical utility of the NICE guideline and the impact of the QOF on diagnosis and management of depression in routine practice’
Slater, 2005 (207)	UK, hospice setting, palliative care	Focus group (n=1)	Nurses (n=4), allied health professional (n=1) support staff (n=3)	Patients and staff completed POS, and feedback was provided to staff	‘The aim of the study was to evaluate the implementation of POS for use in the day hospice setting to improve patient care’

Reference	Location, setting and focus	Study design	Participants	PROMs feedback	Study Aims
Tavabie, 2009 (208)	UK, primary care, depression	Semi-structure interviews and focus groups	GPs (n= 20)	Patients completed PHQ-9 on a computer generating immediate feedback for professionals	‘To identify effects of using mental health questionnaire on views of GPs managing depression, and how this might influence patient care’
Unsworth, 2011 (209)	UK, counselling service, psychological therapy	Focus groups (n=2)	Therapists (n=9)	Patients completed CORE-Net on computer generating immediate feedback for professionals	‘The purpose of this study was to answer the research question: How do National Health Service (NHS) therapists and clients perceive and experience CORE-Net?’
Wressle, 2003 (210)	Sweden, day treatment programme, rheumatoid arthritis	Interviews	Psychotherapists (n=2) Occupational therapists (n=2) Physician (n=1) Social worker (n=1) Assistant nurse (n=1)	Patients completed the COPM and feedback was provided to interdisciplinary team members	‘The aim of this study was to investigate whether the structured method focused on client involvement, the COPM, could work as a tool for a rehabilitation team in a day treatment programme for clients with rheumatoid arthritis’

Study characteristics

Over half of the included studies were carried out in the UK (n=9). The remainder took place in Sweden (n=3), Australia (n=2), the USA (n=1), and Canada (n=1). The study settings included primary care (n=5), hospital care (n=4), hospice care (n=2), and mixed settings (n=4). The setting of one study was not clear (202). The majority of studies were carried out in the context of empirical work (n=12), the remainder were based on the implementation of a national policy (n=4).

The healthcare professionals studied included physicians (n=4), nurses (n=2) and therapists (n=1). Eight studies included a mixture of healthcare professionals and one study did not explicitly state the healthcare professionals involved (202). The treatment focus of the studies was mental health (n=7), palliative care (n=5), oncology (n=1), acute care (n=1), respiratory medicine (n=1) and rheumatoid arthritis (n=1).

Qualitative data was collected through interviews in nine studies, focus groups in five studies and a mixture of interviews and focus groups in two studies. Most studies provided PROMs feedback to healthcare professionals at the individual patient level (n=13). Two studies provided feedback about the average scores of groups of patients and in one study this aspect of the design was unclear (200). All studies provided insights into how PROMs data are used by professionals in practice and a subset of 11 studies also explored the feasibility of data collection.

The quality appraisal exercise found that the included studies were generally good at justifying the research design, providing details on the participants included in the research, explaining the data collection process, clarifying ethical issues, outlining the data analysis methods and the findings, and identifying the value of the research. However, some shortcomings which emerged from the critical appraisal included

unclear rationale for the sampling methods used; a failure to explicitly justify the chosen data collection methods; inadequate incorporation of reflexivity into the research process; insufficient detail about the rigour of analysis; and inadequate methods to increase the credibility of findings (Appendix 3). Three studies were judged to be of a higher standard than the rest on these latter criteria (199, 206, 208).

Synthesis of results

The themes and subthemes which emerged from the thematic synthesis are described in Table 6, and excerpts from the original studies are provided for illustrative purposes. A detailed description of the themes identified in each study is displayed in the Appendix 4. As each paper had slightly different aims, their overall contribution to each theme depended on the focus of the original studies.

Theme 1: Practical considerations

This theme captures issues around the data collection process and the effective use of the information. Practical issues were identified in 14 studies (78, 105-106, 200-210). In nine studies, the workload associated with collecting and analysing data was identified as a significant barrier to the routine use of PROMs (105-106, 200, 202-207). However, some of the studies identified that workloads could be reduced if PROMs feedback was integrated naturally into the consultation process (204, 208-209). The difficulty or ease of PROMs administration also emerged as a determinant of successful implementation. Barriers emerged when the questionnaire was not user-friendly (105-106, 200-203, 205-208), but data collection was facilitated when patients had few difficulties completing the measure (200-202). Some studies identified a lack of collaboration between colleagues as leading to the burden of data collection being placed on a small number of staff members (106, 200, 203, 208). Lack of clear guidelines on the data collection process (patient eligibility, timing,

frequency and location of administration), and on how to correctly analyse and interpret the data created further barriers (105, 200-201, 204-206, 210). However, some studies identified that flexibility in the data collection process was necessary due to variability in the acuity of patients (202, 209). Professionals were more willing to engage in the process when management showed appreciation for the additional work involved and when management themselves became deeply involved in the process (105-106, 200).

Study participants also stated that appropriate training was necessary to effectively engage in the process. They specifically proposed that a lack of training on how to recruit patients, deal with difficult scenarios and effectively use the information created inevitable barriers (105-106, 200, 203-204, 206, 209). Some studies found that having time to become familiar with the measures prior to implementation was a facilitating factor (105-106, 202, 205, 209). Professionals recognised that support during the initiation stage of the data collection was helpful. The effective use of PROMs data was curtailed when statistical support was not available as professionals lacked the expertise to appropriately analyse and interpret the data (106, 200, 206-208). Professionals recognised that they also required support from the wider service to adequately deal with the issues that the measurement highlighted such as referral to specialist professionals or access to suitable treatments (206, 208). Lastly, the use of technology was recognised as a barrier when it slowed down the process (105-106, 209) and a facilitator when it made the collection of the data and dissemination of the findings more efficient (78, 105, 204).

Theme 2: Valuing the data

This theme captures professionals' attitudes to the use of PROMs. It was identified in 11 studies (105-106, 198-199, 203-204, 206-210). Barriers to appreciating the

value of PROMs emerged when the objectives for collection were not transparent. In such circumstances, professionals questioned the motives behind the data collection and expressed fear about how the results would impact on their practice and patient care (105-106, 199, 203, 207, 209). Furthermore, barriers were identified when professionals were not open to receiving feedback or changing their clinical practice (105-106, 198-199, 204, 206-210).

Theme 3: Making sense of the data

This theme captures the methodological considerations that are associated with PROMs. Methodological factors were identified in 13 studies (78, 105-106, 199-200, 202-208, 210). The interpretability of PROMs data influenced professionals' opinions about their scientific value in a quality improvement context (105). Professionals appreciated the graphic presentation of results (204), but identified the need for more sophisticated feedback which clearly depicts what constitutes a clinically important change (105). Others requested aggregated data about the effectiveness of different treatments to complement data about individual patients (78). Concerns about the validity of PROMs emerged in many studies as professionals questioned whether the data produced a genuine reflection of care (105-106, 199, 202-203, 205-208, 210). Professionals identified situations where the validity of measurement was compromised including when patients did not complete the measures accurately, provided socially desirable responses, hid symptoms, failed to follow instructions, or when staff administered the measure incorrectly or in a non-standardised manner. Some professionals also criticised the sensitivity of the measures to accurately detect a change in specific patient populations (200, 202, 207).

Theme 4: Impact on patient care

This theme was identified in all studies and captures issues around the impact of PROMs on care processes and outcomes. There were mixed views regarding the causal link between the use of PROMs and improvements in patient care.

Professionals identified that the use of PROMs in practice had the potential to improve the processes of care by enhancing communication, increasing patient education, promoting joint-decision making, screening for health issues, monitoring changes in disease severity and response to treatment, and stimulating better care planning. Professionals appreciated PROMs as a tool to complement their own clinical judgement and to stimulate professional development. The role of PROMs was also recognised as a research and audit tool (200, 202-203). However, some professionals found that the measures were not of clinical value as the results provided them with no new information (78, 105-106, 198, 200, 202, 205-207). Professionals highlighted some indirect effects of using PROMs on patient care. Negative effects included the intrusive nature of collection on the patient's privacy and the doctor-patient interaction, the capacity to narrow the focus of a consultation, and the opportunity cost for what were perceived to be more important aspects of care. Furthermore, professionals found that certain questions distressed patients and thought the process had the potential to damage the patient-clinician relationship (105-106, 199-200, 202-203, 205-208). Positive indirect effects of collecting PROMs were also identified, which included the ability to build patient confidence in the competence of the professional, to manage patient expectations and to assist in handing responsibility of care back to the patient (78, 199-200, 203, 205, 208-209).

Table 6: Taxonomy of themes, their definitions and excerpts from the studies

Themes	Sub-themes	Definition	Excerpts
Practical considerations	Time/Workload	The impact of PROMs on workloads	<i>Barrier:</i> ‘I think time is the critical issue and that we are being asked to spend more and more time on collecting information and filling out forms’ (105) <i>Facilitator:</i> ‘Some doctors claimed that this intervention might save time, since it provides information in a systematic, time-effective way’ (204)
	Administration	The difficulty or ease of collecting PROMs	<i>Barrier:</i> ‘There were a number of nurses who reported difficulties administering the HQLI. The primary difficulty was patient’s confusion with the answer scales’ (201) <i>Facilitator:</i> ‘Participants reported POS to be easy to use, brief and relevant’ (202)
	Collaboration	The level of cooperation among colleagues	<i>Barrier:</i> ‘I tried to leave [POS] questionnaires for people in the diary and it just didn’t work. I actually came in [on days off] to do it because I rang up to see if anyone had bothered and they hadn’t’ (203)
	Guidelines	The provision of clear or flexible guidelines	<i>Barrier:</i> ‘The hospice ARC (Action Research Collaboration) debated the frequency of POS administration at most meetings’ (200) <i>Facilitator:</i> ‘They expressed the need for user flexibility when using it’ (209)
	Involvement of management/ Use of data	The level of management involvement in the process, and the active use of the information to guide decision making	<i>Barrier:</i> ‘Many staff were frustrated that senior medical staff did not fully appreciate the process’ (106) <i>Facilitator:</i> ‘Senior staff had pre-empted these concerns by discussing POS scores at weekly team meetings so enabling all staff to see the importance and relevance of the data’ (200)
	Training/ Familiarisation	The provision of training and time to become familiar with measures prior to implementation	<i>Barrier:</i> ‘I think we had little education about it really, they’ve just said this is QOF, this is what you’ve got to ask and they’re the questions. We didn’t really have any training’ (206) <i>Facilitator:</i> ‘It was recognized that as one became familiar with the measures the time required for data entry was considerably reduced’ (106)

Themes	Sub-themes	Definition	Excerpts
	Technology	The use of technology for collecting and disseminating the data	<i>Barrier:</i> ‘Access to computers, slowness of the computer networks, lack of computer skills among staff, forgetting passwords and understanding the summary graphs were frequently mentioned’ (106) <i>Facilitator:</i> ‘Allowing the patient to complete the test at home and having the results transferred directly to the doctor’s computer before the consultation’ (204)
	Support	The provision of adequate support to correctly collect, analyse and interpret the data, and support from the wider service to help provide appropriate care	<i>Barrier:</i> ‘This required more statistical analysis than was available to both settings’ (200) <i>Facilitator:</i> ‘There are many things that crop up once you start collecting the data ... it’s great to have someone to call on for help’ (106)
Valuing the data	Transparent objectives	The provision of transparent objectives for collecting PROMs	<i>Barrier:</i> ‘Staff became disappointed in its performance as a patient-assessment tool, the staff’s perception of its purpose became ambiguous, and there was uncertainty as to whether POS was an audit tool by which their effectiveness would be monitored by management’ (207)
	Open to feedback and change	The openness to receiving feedback and willingness to change practice	<i>Barrier:</i> ‘I have my own way of doing things’ (198) <i>Facilitator:</i> ‘The cornerstone of good practice... a type of psychiatric X-ray that shows you where the problems are and how good our treatment... interventions are at sorting out these problems’ (106)
Methodological considerations	Interpretation	The ability to make sense of the feedback	<i>Barrier:</i> ‘Your gut feeling about how depressed someone is and their PHQ-9 score often don’t marry up’ (206) <i>Facilitator:</i> ‘Some clinicians were seeking more sophisticated feedback than just graphs showing current or current-compared-with-past ratings’ (105)
	Validity of measures	The belief that results were a true reflection of care	<i>Barrier:</i> ‘They were also aware of the potential for manipulating scores’ (199).

Themes	Sub-themes	Definition	Excerpts
	Sensitivity	The sensitivity of the measures to detect change	<i>Barrier:</i> 'Direct clinical benefits of using the POS were less apparent to hospice staff, probably owing to the complex clinical needs of their patients that the POS is not sensitive enough to detect ' (200)
Impact on patient care	Quality improvement	The impact of the information on patient care	<i>Barrier:</i> 'QOF tick-box exercise as far as I'm concerned' (206) <i>Facilitator:</i> 'Clients were given the opportunity to identify their own problems, and to make priorities according to what was meaningful to them, this resulted in more distinct goals than before they started to use the COPM '(210)
	Indirect effects	The additional factors that may impact on patient care	<i>Barriers:</i> 'I've actually had people say it, they just make them feel worse... I know how bad I feel and I don't need to see it written down' (205) <i>Facilitator:</i> 'I think that people will develop a respect for your clinical judgement if you spend time listening to them' (208)

Explaining the findings

The relationship between themes and study characteristics was examined to help explain the findings. The characteristics examined included the professional group under study, the study setting, the healthcare issue under examination and the function of the PROM. No explicit pattern was explained by the inclusion of different professionals, settings or healthcare issues. However, the function of the PROMs used in individual studies may have influenced the study findings. Practical facilitators were most likely to be observed in studies where PROMs functioned as a care management tool; however these studies also tended to use computer administration and feedback (78, 105-106, 204, 208-209). A similar trend was observed with the facilitators identified in the methodological theme (78, 105-106, 204). In addition, a lack of clarity regarding the objectives for measurement emerged as a barrier, and involvement of management emerged as a facilitator, when PROMs were used as performance monitoring tools (105-106). Only one study did not identify any positive impacts of using PROMs. This study employed PROMs as a screening and care management tool for mental health issues (206). The studies which did not identify any negative aspects of collecting PROMs employed PROMs as care management tools (201, 204, 209-210).

Risk of bias

The three studies identified as being of a higher quality did not identify any unique themes or sub-themes (199, 206, 208). However, one of these studies exclusively did not identify any positive effects of using PROMs in practice (206).

3.5 Discussion

The barriers and facilitators identified in this review were categorised into practical considerations, attitudes towards the value of the data, methodological concerns, and the impact of feedback on patient care. Practical considerations included workload implications, the ease of data collection, the level of collaboration among colleagues, the provision of clear guidelines for implementation, the level of managerial involvement, the availability of training and support, and the use of technology. Attitudes towards the use of PROMs were associated with the transparency of objectives, and the openness to feedback and change. Methodological concerns identified included the interpretability of the information and the validity of the measures. The impact of the feedback depended on the usefulness of the information to guide decisions on patient care and the indirect effects of routinely collecting PROMs data.

There is a subtle but important distinction between the need for support to correctly analyse and interpret PROMs data, which we have classified as a practical issue, and the concerns raised by professionals about the validity and interpretability of PROMs, which we have classified as a methodological issue. In the ‘practical’ theme, we are addressing the support (statistical help and training) that professionals feel they need in order to familiarise themselves with a relatively alien concept. This is different from fundamental scientific concerns about PROMs which may endure even if statistical support and training are provided.

The themes presented in this review were consistent across different studies. There was some evidence that PROMs were viewed more positively when they functioned as care management tools for individual patients and more negatively when producing performance data about the care delivered by professionals to groups of

patients. This may indicate that PROMs have more value to professionals when they produce data that can be linked to individual patient care, but this interpretation should be considered with caution due to the small number of studies where PROMs were used as performance monitoring tools.

Strengths and limitations

This is the first review to synthesise the qualitative evidence on the experiences of professionals who have first-hand experience of the use of PROMs as a means to improving the quality of healthcare. This review has some limitations. First, the review only focused on English-language articles and it is possible that different experiences with the use of PROMs may be apparent in countries where English is not the first language. Second, only one reviewer performed the initial screening and study selection, and although reference searching was performed to reduce the likelihood of missing appropriate studies there is still a small chance that some relevant literature was missed. Third, the results are based on the credibility of findings in the original studies and there is a lack of detail in all but three studies about the use of methods to enhance credibility. However, the themes identified are quite logical and are similar to those presented in previous reviews of the use of outcome measures generally (191-192). Fourth, the study presents only the perceptions of healthcare professionals and it does not attempt to represent the views of patients or healthcare managers about the value of PROMs. Fifth, the study did not attempt to explore professionals' perceptions of PROMs feedback in the context of empirical work versus as a result of a policy implementation. This may be an important focus for further research as motivations for change and attitudes may differ. In the former, change may be promoted from the bottom-up compared to the latter where change may be promoted from the top-down.

Relevance to previous literature

The themes identified in this systematic review are well-known barriers and facilitators to the success of audit and feedback interventions in other contexts. Our systematic review confirms the importance of these issues while revealing new insights specific to PROMs. For example, practical barriers such as inadequate organisational and technical support have been comprehensively documented in the quality improvement literature (211-213). This review deepens our understanding of these issues in the context of PROMs by highlighting the considerable barriers associated with data collection, and the need for specific training in the use and interpretation of psychometric instruments. Similarly, there is evidence from the broader literature that interventions are more likely to fail when professionals display negative attitudes and are suspicious about the purpose of audit and feedback (214-216). Our review highlights the specific issues associated with negative attitudes to PROMs, including methodological concerns about the validity of patient-reported data and worries about the potential for routine PROMs administration to disrupt patient care. It is of note that these concerns have also been voiced by patients in separate qualitative studies (217-218). Finally, there is evidence from other contexts that feedback has the greatest impact when it is focused on specific task based solutions and delivered in a goal-setting context (30, 219). Our review underlines how difficult it is for PROMs to satisfy these criteria given the problems experienced by professionals in attempting to interpret PROMs feedback and turn the information into concrete quality improvement solutions.

Implications for clinicians and policymakers, and future research

It is clear that many professionals remain to be convinced about the value of PROMs but that they could be encouraged to engage with their use given the right practical

and methodological support. Greater investment in data collection technology could relieve much of the human workload and make feedback more timely (220). Greater clarity over the objectives of data collection and investment in methodological training are additional solutions. It is interesting that PROMs feedback have shown greatest promise in the area of mental health, a field where the use of these measures has long been embedded in routine practice, and where professional attitudes may be more positive as a consequence (141, 174-175, 221). However, it is important to understand the cause of any resistance as professionals may have good reasons for not implementing or using PROMs (222). For example, PROMs have well known problems with interpretability and professionals may therefore have legitimate grounds for resisting their use (190, 223). The appropriateness of using PROMs in a quality improvement context is also a source of legitimate debate. Most commonly used PROMs were developed to evaluate the effectiveness of different treatments and therefore may not provide sufficient or appropriate information to guide quality improvement activities. This problem is indicative of a relatively poor theoretical basis for the use of PROMs in a quality improvement context (147).

The barriers identified in this review may represent a failing on the part of those who advocate the use of PROMs to sufficiently engage professionals in the planning stage and to acknowledge the conflict between managerial and professional objectives (149, 224). A deeper understanding of the motivations of different stakeholders is essential to disentangle how PROMs can be used to improve quality in reality. Further qualitative studies with professionals and case-studies of PROMs initiatives are essential (104). This would help researchers and policy makers gain an understanding of how this information impacts on clinical decision making. Lastly, evidence is required to identify the specific healthcare issues and patient populations

that have large variability in outcomes as these are where PROMs data is likely to have the greatest impact. Otherwise, as Wolpert points out, inappropriately implementing PROMs in practice may only lead to an increased bureaucratic burden with little positive impact on care (225).

**The Patient-Reported Outcomes: Feedback Interpretation
and Learning Experiment Trial-
The PROFILE Trial**

This article is being prepared for publication.

Chapter 4: The PROFILE Trial

4.1 Abstract

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

Design: A cluster randomised controlled trial with a repeated cross-sectional design and a six-month follow-up.

Setting: Secondary care in the Republic of Ireland.

Participants: Surgeons were recruited through the Irish Institute of Trauma and Orthopaedic Surgery and patients were recruited either in a pre-assessment clinic or on the wards prior to surgery. We randomly allocated 21 surgeons and 592 patients to intervention or control groups.

Intervention: Surgeons in the intervention group received peer benchmarked PROMs feedback and an educational session. Surgeons in the control group did not receive feedback or education.

Main outcome variable: Post-operative Oxford Hip Score (OHS).

Results: Primary outcome data were available for 11 intervention surgeons with 215 patients and for 10 control surgeons with 217 patients. The mean post-operative OHS for the intervention group was 41.1 (95% CI 40.1-42.0) and for the control group was 41.9 (95% CI 41.0-42.7). The adjusted effect estimate was -0.7 (95% CI -1.9-0.5, P=0.2). Secondary outcomes were the Hip Osteoarthritis Outcome Score (HOOS), EQ-5D and the proportion of patient reporting a problem after surgery. The mean post-operative HOOS for the intervention group was 36.5 and for the control

group was 37.2. The adjusted effect estimate was -0.2 (95% CI -1.5-1.2, P= 0.8). The mean post-operative EQ-5D for the intervention group was 0.85 and for the control group was 0.87. The adjusted effect estimate was -0.003 (95% CI -0.03-0.02, P=0.7). 27% of patients in the intervention arm and 24% of patients in the control arm reported at least one complication after surgery. The adjusted odds ratio was 1.1 (95% CI 0.6-2.4, P=0.6).

Conclusions: Providing surgeons with peer benchmarked PROMs data did not result in better patient outcomes compared to outcomes of surgeons who did not receive feedback. Currently, we do not fully understand the extent to which outcomes data promotes providers to improve quality. PROMs information alone tends to be insufficient to help identify opportunities for quality improvements. In addition, conceptual, methodological practical and attitudinal issues are common as surgeons have problems with understanding the unique nature of PROMs, they have methodological concerns about PROMs particularly with respect to the validity and interpretation of the data, and they frequently encounter practical issues such as workload pressures and a lack of support preventing the appropriate use of the information.

4.2 Introduction

Patient-reported outcome measures (PROMs) are questionnaires that assess patients' health, health-related quality of life and other health-related constructs (226). 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 (3, 72-73, 103-108). The NHS PROMs Programme in England is the most advanced example of this approach (3). Introduced in 2009, it mandates the collection of PROMs for all patients undergoing hip replacement, knee replacement, hernia repair and varicose vein surgery (1). The data, which are publically reported online at the NHS Trust level (87), can be used by patients and purchasers to select an NHS Trust for their surgical procedure, and by Trusts to stimulate quality improvements (3, 100).

The evidence to support the use of PROMs as performance measurement tools is weak (141). In 2013 a systematic review of randomised controlled trials found only one study which evaluated the impact of providing peer benchmarked PROMs feedback to primary care physicians. This study focused on the functional status of 1116 elderly patients under the care of 48 physicians and found no intervention effect. In fact, patients in both the control and intervention arms had a statistically significant decrease in functional status over the course of the study (179). A subsequent time-series analysis evaluated the impact of the NHS PROMs Programme over the period 2009 to 2012 and found no consistent positive effect on patient outcomes. The study authors concluded that the lack of impact could be explained by an inadequate implementation strategy (227). However, it is important to also consider the possibility that the results of this observational study were biased

by time varying confounders such as changes in resources, workforce composition and technology over the study period (227).

With respect to implementation, peer benchmarking may be more valuable if it is provided to individual clinicians as opposed to organisations (100). In addition, a 2014 systematic review of the qualitative literature on PROMs feedback suggested that the information may be more useful if it is delivered in a clear format and is supported by education on the interpretation of results (142). There is also the question of whether the measures being used to measure performance are fit for purpose as they have not been developed and psychometrically tested for this use (228).

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.

4.3 Methods

This study was a cluster randomised controlled trial with a 1:1 allocation ratio of surgeons to an intervention or control arm. We tested the hypothesis that providing peer benchmarked PROMs feedback and educational support to individual surgeons would result in better patient outcomes.

As the intervention was designed to improve the outcomes of patients by enhancing the performance of healthcare professionals, a cluster randomised controlled trial was used (229). Eligible professionals 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 included if they were under the care of eligible surgeons, over 18 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 (1).

Intervention

The feedback intervention was designed to replicate the methods used in the NHS PROMs Programme, with the exception that feedback was provided to individual surgeons in the intervention group rather than NHS Trusts, and feedback was accompanied with educational support. The feedback report was designed using the results of a qualitative study which explored professionals' preferences for metrics used to compare performance (230). Each surgeon was provided with feedback derived from a PROM, the Oxford Hip Score (OHS) (231). When drawing statistical comparisons of surgeons' performance, case-mix adjustment of the OHS was undertaken to account for patients' pre-operative OHS, age, sex, general health status and mental health status (232). Surgeons were also provided with feedback on the proportion of patients that reported an overall improvement in their hip problem and the proportion of patients that reported having at least one of four problems after surgery for patients under their care. Statistical comparison of surgeons' performance on these metrics were unadjusted for case-mix following methods used in the NHS PROMs Programme (232). The report presented to individual surgeons clearly demonstrated how each surgeon performed in comparison to the other surgeons in the trial; however the identity of the other surgeons remained anonymous (Appendix 5). The report was delivered to surgeons in the intervention group in

January 2013 by post and email. In addition, an educational video session was produced by an expert on the interpretation of PROMs data (JB), and was made available to surgeons in the intervention arm by an email link to a dedicated website. The educational session described the outcome measures and explained the correct interpretation of the graphs included in the report, such as how to identify statistically significant and clinically important differences (Appendix 6). 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 the feedback was the mean difference in post-operative OHS for patients operated upon after feedback was delivered to the intervention group of surgeons. The OHS is the disease-specific measure used by the NHS PROMs Programme (1). It consists of 12 items on symptoms and functional status with five levels of response. Each item can score 0-4 summated to an overall score of 0 (worse health status) to 48 (best health status) (231).

Secondary outcome measures included a version of another disease-specific measure— the Hip Osteoarthritis and Outcome Score (HOOS) (233), a generic quality of life measure— the EQ-5D (234), and the proportion of patients reporting an allergy or reaction to a drug, urinary problems, bleeding or wound problems after surgery. The HOOS consists of 11 items on symptoms and functional status with five levels of response. Each item can score 0-4 summated to an overall score of 0 (worse health status) to 44 (best health status). The EQ-5D is based on five dimensions (mobility, self-care, daily activities, pain and anxiety/depression) with

three levels of response. Utility scores are generated using a standard algorithm and range from -0.59 (worse than dead), 0 (dead) to 1 (perfect health)(234).

To deal with missing items, the mean response of the items that the patient completed were imputed if they had not missed more than five questions for the OHS and the HOOS, and the mode response of the items that the patient completed were imputed if they had not missed more than two questions for the EQ-5D (1).

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 surgeons that may have been potentially eligible for the trial. These additional surgeons were contacted by phone and five agreed to participate. The 35 consenting surgeons represented 95% of the 37 high volume hip replacement surgeons in the Republic of Ireland at the time of recruitment.

Data collection occurred in two phases, before and after randomisation to the feedback and control arms of the trial, and a different cohort of patients was recruited in each phase. The cohort of patients used to generate feedback to surgeons was recruited over the period May 2011 to June 2012. The cohort of patients used to assess the effectiveness of the feedback was recruited over the period February 2013 to December 2013.

During both data collection phases, nurses and registrars identified and recruited eligible patients prior to their operation in a pre-operative assessment clinic, if available, or alternatively when patients were admitted for surgery. MB provided training to the data collectors at each site to standardise procedures (Appendix 7).

Patients in both the pre-feedback and post-feedback phases 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 (Appendix 8). Data collectors were asked to complete a 'Patient Participation Form' to enable us to account for all eligible patients; it detailed patients who were invited to participate in the study, patients who were excluded due to ineligibility, patients who refused to consent and lastly, patients who were missed (Appendix 9).

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 six months after their operation. Post-operative data collection was managed by MB. Questionnaires were posted to patients and a reminder was sent four weeks later if a reply was not received within this timeframe. Pre-operative questionnaires included demographic questions on the patient's age, sex and duration of symptoms, the OHS, a version of the HOOS, the EQ-5D and a general health status item (Appendix 10). Post-operative questionnaires included the same questions as the pre-operative questionnaire plus questions on the results of the operation (Appendix 11).

Sample size

This trial required separate sample size calculations for the pre-feedback and post-feedback phases of the study. The first calculation established the number of patients required to accurately benchmark surgeons for the feedback intervention. 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 (235) between the average score for one surgeon and the average score for all surgeons, with 80% power at the 5% significance level. We inflated this to 32 patients per

surgeon to allow for attrition during post-operative follow-up. This was set as the minimum recruitment target for each surgeon during the pre-feedback phase of the trial. The second calculation established the sample size to detect a significant difference in outcome between patients in the feedback and control arms of the trial after randomisation. Using data collected during the pre-feedback phase we identified that the extent of within surgeon clustering of the post-operative OHS, as measured by the intra-class correlation coefficient (ICC), was 0.03. As 21 surgeons achieved the recruitment target in the pre-feedback phase, the number of clusters was fixed at 21. Therefore, to detect a difference of four points in the OHS between the feedback and control arms of the trial with 90% power at the 5% significance level, we calculated that data on the primary outcome for 114 patients would be necessary in each arm. We inflated the recruitment target to 148 patients for each study arm to allow for a loss of 30% of patients to follow-up, thus giving a total sample size target of 296 patients for the post-feedback phase of the study.

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 21 surgeons who achieved the target recruitment in the pre-feedback phase. Surgeons were stratified according to public/private status of the hospitals within which they practiced and whether their performance, as measured by the OHS, during the pre-feedback phase of the trial was above or below average (236). 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 Stattrek program.

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 and those recruiting patients were unaware of the trial arm to which surgeons had been allocated throughout the study.

Statistical Analysis

A linear mixed effects regression model was used to evaluate the effect of PROMs feedback on the primary outcome between 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 post-operative OHS. In the main analysis, we used data from all patients who had post-operative data and we adjusted the effect of PROMs feedback for the influence of patient level characteristics (age, sex, pre-operative score and general health status). Similar methods were used to evaluate the effect of PROMs feedback on the secondary outcomes. A linear mixed effects regression model was 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-responders, pre-operative characteristics of patients who did not respond were compared across arms. We also carried out a sensitivity analysis by imputing the last observation carried forward for patients lost to follow-up. Two additional sensitivity analyses were carried out to assess the impact of imputing missing items on the estimate of the effect of feedback and to examine the impact of including the hospital identity as a random effect into the mixed effects model.

To test the change in outcomes across arms from the pre-feedback phase to the post-feedback phase, we used linear and logistic regression models which were adjusted

for patient characteristics. For all tests, we used a value of 0.05 for the level of significance. The results report means and odds ratios with 95% confidence intervals (CIs).

Ethics

This study was conducted according to ethical guidelines. The Research Ethics Committee of the Cork Teaching Hospitals approved the study protocol, as well as individual ethics committees within participating hospitals.

4.4 Results

Overall, 21 surgeons achieved sufficient patient recruitment in the pre-feedback phase to be included in the trial. Eleven were randomised to the intervention arm and ten to the control arm. All participating surgeons were male and had been consultants for ten years on average. Nine surgeons worked in a public hospital only, four worked in private hospitals only, and nine worked in both public and private hospitals. Surgeon characteristics were similar across the study arms (Table 7).

The reports provided to surgeons in the feedback arm of the trial contained information on an average of 27 patients per surgeon (range 22-42). The reports covered 312 patients in the feedback arm and 261 patients in the control arm. The mean adjusted change in OHS for all patients in the pre-feedback phase was 21.5 (95% CI 20.8-22.0): the figure was the same for patients of surgeons who were eventually allocated to the control arm 21.5 (95% CI 20.6-22.3) and the intervention arm 21.5 (95% CI 20.6-22.3). A response rate of 82% was achieved in the pre-feedback phase of the trial for patients of surgeons who were eventually allocated to the feedback and control arms. Patients excluded due to non-response tended to be younger and have worse pre-operative scores than those included in the study.

Patients excluded due to surgeons not reaching sufficient recruitment levels reported slightly worse post-operative scores than those included in the study (Appendix 12).

The characteristics of patients of surgeons who were eventually allocated to the intervention and control arms were similar (Appendix 13).

For the post-feedback phase of the trial, 592 patients were recruited across the 21 surgeons. 288 patients were under the care of surgeons in the intervention arm and 304 patients were under the care of surgeons in the control arm (Figure 7). We estimated that 51% of patients from surgeons in the intervention arm and 58% of patients from surgeons in the control arm were considered for participation in the study. Of these 2% of the patients in both arms were considered ineligible for participation and 7% were invited to participate but refused to consent. Patient characteristics were similar across arms, except for patients in the intervention arm having slightly worse pre-operative EQ-5D scores than those in the control arm (Table 7). A response rate of 78% was achieved for the intervention group and 84% for the control group. Patients lost to follow-up in the intervention arm reported slightly worse pre-operative EQ-5D scores than those lost to follow-up in the control arm. 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 for each surgeon was 38 weeks for surgeons in the intervention arm (range 19-49) and 36 for surgeons in the control arm (range 17-49).

The total number of patients recruited for the post-feedback phase was greater than our sample size target as the attrition rate was lower than estimated and some surgeons recruited more patients than expected over the study period.

Primary outcome

Table 8 presents the effect of the intervention on outcomes. The unadjusted mean post-operative OHS for all patients was 41.5 (95% CI 49.8-42.1). The unadjusted mean post-operative OHS for the intervention group was 41.1 (95% CI 40.1-42.0) and for the control group was 41.9 (95% CI 41.0-42.7). After adjusting for patient characteristics, the mean difference was -0.7 (95% CI -1.9-0.5, P=0.2). The adjusted mean difference obtained from the linear mixed effects model was also -0.7 (95% CI -1.9-0.5, P=0.2). The adjusted mean difference in the post-operative OHS between the pre-feedback phase and post-feedback phase of the study for the intervention group was 0.3 (95% CI -0.9-1.5, P=0.6) and for the control group was 0.9 (95% CI -0.1-2.0, P=0.08).

Secondary outcomes

Table 8 also presents the effect of PROMs feedback on the secondary outcomes. The unadjusted mean post-operative HOOS for all patients was 36.8 (95% CI 36.1-37.5). The mean post-operative HOOS for the intervention group was 36.5 (95% CI 35.5-37.5) and for the control group was 37.1 (95% CI 36.2-37.9). The adjusted effect estimate obtained from the linear mixed effects model was -0.2 (95% CI -1.5-1.1, P=0.8). The adjusted mean difference in the post-operative HOOS between the pre-feedback phase and post-feedback phase of the study for the intervention group was 0.5 (95% CI -0.8-1.8, P=0.4) and for the control group was 0.9 (95% CI -0.3-2.0, P=0.1).

The unadjusted mean post-operative EQ-5D for all patients was 0.86 (95% CI 0.85-0.88). The mean post-operative EQ-5D for the intervention group was 0.85 (95% CI 0.82-0.88) and for the control group was 0.87 (95% CI 0.85-0.89). The adjusted effect estimate obtained from the linear mixed effects model was -0.003 (95% CI -

0.03-0.02, P=0.7). The adjusted mean difference in the post-operative EQ-5D between the pre-feedback phase and post-feedback phase of the study for the intervention group was 0.01 (95% CI -0.01-0.04, P=0.2) and for the control group was 0.02 (95% CI -0.00-0.05, P=0.05).

The unadjusted percentage of all patients that reported a problem after surgery was 25% (95% CI 21-30). The percentage of patients that reported at least one complication after surgery in the intervention arm was 27% (95% CI 21-33) and in the control arm was 24% (95% CI 18-30). The adjusted effect estimate obtained from the logistic mixed effects model was 1.1 (95% CI 0.6 -2.4, P=0.6). The adjusted odds ratio of patients reporting a problem in the pre-feedback compared to post-feedback phases of the study was 0.89 (95% CI 0.6-1.5, P=0.9) for the intervention group and 0.99 (95% CI 0.6-1.5, P=0.9) for the control group.

Sensitivity analysis

When comparing the post-operative 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 -0.7 (95% CI -1.9-0.5, P=0.2), when imputing values for patients that were lost to follow-up -0.7 (95% CI -1.9-0.6, P= 0.2) and when the hospital identifier was included as a random effect -0.7 (95% CI -1.9-0.5, P= 0.2). Similar results were found for the HOOS when the missing items were not imputed -0.1 (95% CI -1.4-1.15, P=0.8), when imputing values for patients that were lost to follow-up -0.1 (95% CI -1.3-1.3, P= 0.9) and when the hospital identifier was included as a random effect -0.1 (95% CI -1.4-1.1, P= 0.8). In addition, similar results were also found for the EQ-D5 when the missing items were not imputed -0.005 (95% CI -0.03-0.02, P=0.7), when imputing values for patients

that were lost to follow-up -0.03 (95% CI $-0.07-0.00$, $P= 0.09$) and when the hospital identifier was included as a random effect -0.005 (95% CI $-0.03-0.02$, $P= 0.7$).

Figure 7: Flow of participants through the study

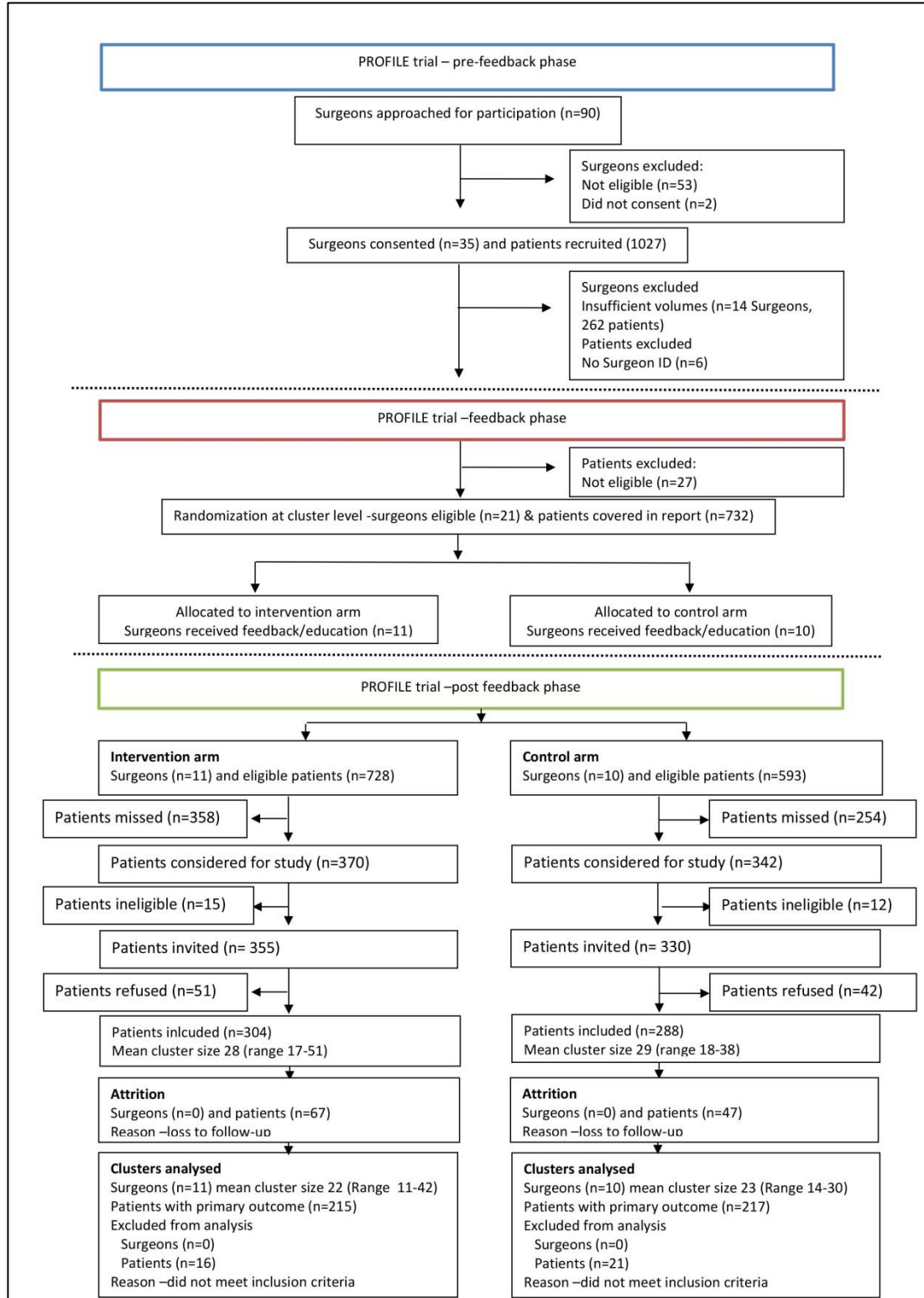


Table 7: Baseline characteristics between arms

Characteristics (<i>Level</i>)	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)	21.5 (1.3)	21.5 (1.5)
Patients covered by feedback report, mean (SD)	27 (4.6)	29 (6.9)
Patient	N=288	N=304
Age, mean (SD)	66.2 (11.2)	64.6 (11.8)
Male, n (%)	141 (53%)	148 (51%)
Health status, n (%)		
Excellent	32 (12%)	30 (11%)
V. Good	90 (34%)	97 (34%)
Good	106 (41%)	126 (45%)
Fair	29 (11%)	17 (6%)
Poor	4 (2%)	11 (4%)
Duration of symptoms, n (%)		
<1 year	43 (16%)	53 (18%)
1-5 years	184 (69%)	187 (65%)
6-10 years	25 (9%)	30 (10%)
>10 years	15 (6%)	18 (6%)
OHS pre-op, mean (SD)	19.9 (8.3)	19.1 (8.5)
HOOS pre-op, mean (SD)	17.7 (7.5)	17.1 (7.8)
EQ5D pre-op, mean (SD)	0.43 (0.31)	0.38 (0.33)*

*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.

Table 8: Primary and secondary outcome (baseline and 6 months)

Outcome and period	Control group		Intervention group		No of patients in multi-variate analysis	Adjusted effect estimate* (Intervention versus control) (95% CI)	P-value	ICC
	No. of patients	Mean (SD)	No. of patients	Mean (SD)				
Primary outcome								
<i>Oxford Hip Score</i>								
Baseline	267	19.8 (8.3)	286	19.0 (8.5)				
6 months	217	41.9 (6.3)	215	41.1 (7.2)	339	-0.7* (-1.9- 0.5)	0.2	0.03
Secondary outcomes								
<i>HOOS</i>								
Baseline	262	17.7 (7.5)	283	17.1 (7.8)				
6 months	214	37.2 (6.4)	215	36.5 (7.3)	336	-0.2* (-1.5-1.2)	0.8	0.03
<i>EQ5D</i>								
Baseline	254	0.43 (0.3)	273	0.38 (0.3)				
6 months	211	0.87 (0.2)	204	0.85 (0.2)	316	-0.003* (-0.03-0.02)	0.7	0.03
<i>Proportion reporting problems after surgery</i>	218	0.24 (0.43)	215	0.27 (0.44)	341	1.1** (0.6-2.4)	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.

4.5 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 significant difference in outcomes for patients treated by surgeons who were randomised to a feedback group compared to patients treated by surgeons who were unaware of their performance. Outcomes for patients in both groups improved slightly over the course of the trial, although the differences between pre-feedback and post-feedback outcomes were not statistically significant.

Explanation of findings

A separately published qualitative study was undertaken to explore the views of surgeons about the value of peer benchmarked PROMs feedback (143). The findings of this study help to explain the apparent ineffectiveness of the feedback intervention. Surgeons had mixed opinions on the value of peer benchmarked PROMs data. Many appreciated the feedback as it reassured them that their practice was similar to their peers. However, PROMs information alone was considered insufficient to help identify opportunities for quality improvements. Three reasons for the observed reluctance of surgeons to embrace PROMs were identified. First, the surgeons had problems with understanding the unique nature of PROMs, for example confusing them with patient satisfaction measures. Second, some surgeons had methodological concerns about PROMs, particularly with respect to the validity and interpretation of the data. Third, practical constraints such as workload pressures and a lack of support were barriers to the uptake of PROMs (143).

One explanation for the slight improvement in both arms of the trial is a Hawthorne effect whereby the performance of surgeons improved through the act of monitoring

alone (101). The second explanation is that there may have been a contamination effect across surgeons from the intervention group to the control group. Although feedback reports were individually tailored for each surgeon in the intervention group, the information may have promoted discussion and debate within hospitals stimulating improvements at the level of the orthopaedic unit (161). Eight hospitals had surgeons that were randomised to both intervention and control arms.

Implications of findings on policy and practice

Performance monitoring can provide information about how professionals perform relative to their peers but it does not explain why performance differs. In theory, the process of peer benchmarking assumes that professionals will be promoted to undertake additional audit or research activities to identify the reasons for differences in performance (2-3). However, this did not happen in practice (143). It is important to identify the capabilities, opportunities and motivations which are linked to the ability and desire to change (51). The incentive to undertake additional audit and research to identify what change in processes of care is required to improve patient outcomes relies on the assumption 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 (227). For example, statistical and analytical support may be necessary to link PROMs data to processes of care measures such as clinical data and patient experience data. Opportunities to improve may be enhanced if the healthcare culture promotes and incentivises continuous quality improvement (41). Improvements often require a level of flexibility within the system to allow changes to patient pathways or to support additional investments in training, equipment and infrastructure.

Finally, professionals may be motivated to change if they receive continuous and timely feedback as this would reinforce the findings and also allow providers to track the impact of any changes to care processes overtime (152). Also, greater motivation to improve may be encouraged if there are consequences for those performing poorly (153, 237).

A good measure of performance should be fit for purpose. Many of the measures commonly used have been developed within the Classic Test Theory paradigm and therefore are limited to the degree to which they are scientifically robust (223). It is important to appreciate that using a PROM with established psychometric properties does not guarantee a reliable and valid performance measure (228, 238). PROMs were not designed to assess the performance of healthcare professionals and as a result may not meet the standards required of such measures. For example, measures such as the OHS and EQ-5D have not undergone the National Quality Forum endorsement for performance measures (228). Another important consideration when planning performance monitoring is that there should be sufficient variation in outcomes between providers to justify the measurement. This concept is related to the ICC. Some believe that performance monitoring is warranted when the ICC is greater than 0.10 (239). The variation between surgeons in this study was small. A low ICC may suggest that the standard of care is already high leaving little room for improvement or that all providers were performing equally poorly. The latter is unlikely as the baseline scores in this study were better than the national average scores observed in a similar cohort in England (1). The lack of impact observed in this study may also be a result of excluding low volume surgeons who may have benefited more from the feedback. We found that surgeons excluded from the trial due to low patient volumes had slightly worse post-operative scores than those

included in the study. This hypothesis is in tandem with a growing body of evidence which demonstrates a positive association between higher procedure volumes and better outcomes following surgery (240).

Strengths and weaknesses of this study

A major strength of this study is the large amount of data that were collected to successfully deliver the feedback intervention. The study employed a complex multi-phase, multi-centre 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 multi-level modelling to account for a lack of independence between observations and surgeon-level effects. The possibility of performance and detection bias was unlikely as data collectors and patients were blind to the allocation of the surgeon. All surgeons who were randomised remained in the trial and did not crossover, and furthermore, patient response rates were high and were 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 (143). Finally, the study included 35 out of a possible 37 high volume surgeons in the country and the model of care for hip replacement surgery in Ireland is similar to models used in other developed world countries, thus the external validity of findings is strong.

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 that were not recruited differed across the intervention and control arms of the study but there is no obvious reason why this should be so. The

recruitment levels observed in this study are similar to those observed in the NHS PROMs Programme (241) and reflect the considerable practical challenges involved in collecting PROMs on a routine basis across different treatment sites (71, 242). 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 was on average 38 weeks. This may not have allowed sufficient time to capture the impact of potential improvements to care structures and processes on patient outcomes. Furthermore, the research is based on only one round of feedback. Professionals may be more likely to engage with using PROMs data if they receive regular feedback reports and can observe consistent trends over time (30). Finally, this research does not explore the influence of feedback on the wider healthcare system and does not investigate the impact of extrinsic forces or motivations which can be employed to improve care such as public reporting or pay for performance (100).

Generalizability of the findings

The context in which this trial was undertaken should be considered to establish the generalizability of findings to other settings. The trial was undertaken in Ireland where routine outcome measurement is not performed so this was the first time surgeons had received peer benchmarked feedback. The lack of impact may be explained by the surgeons' unfamiliarity with the use of PROMs as a performance measurement tool (215). Feedback was provided at the surgeon level; however the NHS PROMs Programme provides feedback at the Trust level. This may lead to different motivations for improvement. Providing anonymous benchmarked feedback relies on surgeon's intrinsic willingness to change rather than institutional forces linked to clinical governance or economic forces linked to public reporting

and payment for performance (2-3, 100). Evidence from the wider literature on the use of performance monitoring has demonstrated that units respond better than individuals to this type of feedback (153, 243). The results should be also interpreted in the context of hip replacement surgery for which we found little variation between surgeons. The impact of performance feedback may be greater for procedures with greater variability between providers (239). Finally, the mechanisms of change in behaviour may differ across professional groups as the ability and flexibility to change care processes of care may vary (100).

Comparison with previous studies

The findings of this study are consistent with a time-series analysis on impact of the NHS PROMs Programme (227) and a randomised study on the impact of providing peer benchmarked PROMs feedback to primary care physicians (179). One of the longest running initiatives that provides peer benchmarked outcomes data to surgeons is the publication of cardiac mortality report-cards in New York State. An initial evaluation, published almost 20 years ago, claimed a positive association between the feedback and outcomes (244). However, since then there has been substantial debate over the effectiveness of the feedback. Many argue that the improvement is attributed to developments in science, as the evidence regarding the extent to which outcomes data promotes providers to improve the quality of care remains unresolved (67-68).

Future research

Although our understanding of how PROMs may impact on behaviour is in its infancy, there are some interesting innovations occurring in this field. The on-going debate about the appropriate use of outcomes data stems from our lack of understanding about the factors which predict or affect outcomes (59). Analytical

approaches used in production economics may help to uncover meaningful relationships such as data envelopment analysis, stochastic frontier analysis and multi-level multivariate modelling (3). Coupled with this potential are qualitative approaches using realist analysis which aim to explain the mechanisms of change that are linked to different quality improvement strategies (100). Furthermore, there are huge efforts being placed on trying to improve measurement. Traditional approaches used to develop measurement scales have been superseded by more sophisticated techniques which have the ability to develop more accurate and efficient measures (223, 228). Finally, advances in technology will generate the capacity to collect and feedback this information in real time (73).

Conclusion

The evidence supporting the use of PROMs as peer benchmarking tools is weak. Efforts to improve quality may involve structural changes to care, the impact of which may take a number of years to filter through to patient outcomes. The implications of this are that the use of PROMs as a performance measurement tool may be incorrectly labelled as being ineffective, when in fact the full effect of the intervention is not captured within the timeframe of a research project. It is also possible that the variation in outcomes between surgeons is too small to justify performance monitoring. Furthermore, in a health system where outcomes are not routinely monitored and clinicians are not familiar with the use of PROMs as a performance measurement tool, a period to allow for sufficient training, education and adaptation may be required before the information is used to promote change.

5

Surgeon's experiences of receiving peer benchmarked feedback using patient-reported outcome measures.

A qualitative study

This article was published as:

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.

Chapter 5- Qualitative Study

5.1 Abstract

Background: The use of patient-reported outcome measures (PROMs) to provide healthcare professionals with peer benchmarked feedback is growing. However, there is little evidence on the opinions of professionals on the value of this information in practice. The purpose of this research is to explore surgeon's experiences of receiving peer benchmarked PROMs feedback and to examine whether this information led to changes in their practice.

Methods: This qualitative research employed a Framework approach. Semi-structured interviews were undertaken with surgeons who received peer benchmarked PROMs feedback. The participants included eleven consultant orthopaedic surgeons in the Republic of Ireland.

Results: Five themes were identified: conceptual, methodological, practical, attitudinal, and impact. A typology was developed based on the attitudinal and impact themes from which three distinct groups emerged. 'Advocates' had positive attitudes towards PROMs and confirmed that the information promoted a self-reflective process. 'Converts' were uncertain about the value of PROMs, which reduced their inclination to use the data. 'Sceptics' had negative attitudes towards PROMs and claimed that the information had no impact on their behaviour. The conceptual, methodological and practical factors were linked to the typology.

Conclusion: Surgeons had mixed opinions on the value of peer benchmarked PROMs data. Many appreciated the feedback as it reassured them that their practice was similar to their peers. However, PROMs information alone was considered

insufficient to help identify opportunities for quality improvements. The reasons for the observed reluctance of participants to embrace PROMs can be categorised into conceptual, methodological, and practical factors. Policy makers and researchers need to increase professionals' awareness of the numerous purposes and potential benefits of using PROMs, challenge the current methods to measure performance using PROMs, and reduce the burden of data collection and information dissemination on routine practice.

5.2 Background

Patient reported outcome measures (PROMs) are questionnaires that assess patients' views about their health (1, 3). They have traditionally been used to assess the burden of disease and to evaluate the clinical effectiveness of different treatments (3). More recently, they have been used to give feedback to healthcare professionals in the hope that such information will lead to improvements in the delivery of care (141).

PROMs feedback can be based on data about individual patients or groups of patients defined at the level of the healthcare provider. Feedback about PROMs for individual patients is intended to help healthcare professionals identify new healthcare issues, assist in monitoring disease severity, and assess the effectiveness of current treatments (82, 147, 245). For example, in an attempt to promote the effective management of chronic obstructive pulmonary disease symptoms patients were asked to complete the Short Form 36 Health Survey on a touch screen computer prior to their consultation, and feedback about the patient's self-reported physical and mental health was provided to the physician during the consultation (246). Feedback about PROMs for groups of patients seeks to stimulate

professionals to consider their performance in comparison to their peers, empower purchasers and patients to select providers on the basis of performance, and facilitate reward mechanisms such as payment by performance (3, 141). For example, the NHS in England introduced the PROMs Programme in 2009 that mandated the collection of PROMs for patients undergoing four common elective procedures (hip replacement, knee replacement, hernia repair, and varicose vein surgery). Patients are invited to complete a disease-specific and a generic measure prior to and after their surgery. Patient data are aggregated to the level of the provider to compare performance and the results are publically reported online at NHS Trust level (2-3, 247).

PROMs have been adopted as quality improvement tools in the UK (3, 72), America (103-104), Australia (105-107), and Sweden (104). In addition, Canada (73) and the Netherlands (108) have imminent plans to implement PROMs into healthcare policy. Arguably, the UK is revolutionising this field by firmly developing a role for PROMs in managing performance (248).

Despite the growing interest in PROMs, a number of systematic reviews have found weak evidence to support their effectiveness in promoting quality improvements (83, 109-112, 141, 145, 155). A recent systematic review of 16 studies examined the impact on patient outcomes of feeding back PROMs data to healthcare professionals. The review found inconclusive evidence of the effectiveness of PROMs feedback about individual patients. Only one study examined the effectiveness of peer benchmarking using PROMs data. This study found no statistically significant difference in patient outcomes between the feedback and control arms (141). In addition, a recent review of the qualitative literature found 14 studies that had explored professional's views on the value of receiving PROMs feedback about

individual patients (142). A further two studies had examined the value of PROMs feedback at both the individual and aggregated level, but it was not possible to separate the results for these different forms of feedback (142). Given that the use of PROMs at the aggregate level presents potentially unique challenges, it is important to examine professional's views and experiences about this specific form of feedback (249). For example, aggregated PROMs data may prove more difficult to interpret than PROMs data about individual patients, and peer benchmarking may be mistrusted because the methods used to perform case-mix adjustment of PROMs data are not widely understood (249-250). These issues may engender confusion and scepticism among those tasked with using the data for quality improvement purposes (228).

This study explores professional's experiences of using PROMs as peer benchmarking tools. The objectives of this research were to identify the practical challenges of collecting and using PROMs data in practice, methodological challenges associated with generating useful PROMs feedback, attitudes towards the value of this feedback, and the impact of this information on stimulating changes to clinical practice and on promoting professionals to undertake additional audit or research activities. This research is timely considering the current plans to expand the NHS PROMs Programme to different conditions and to begin publishing data at the individual consultant level (3, 248).

5.3 Methods

Design overview

This paper reports on a qualitative research study that was nested within a larger randomised controlled trial of PROMs feedback. The trial was titled the Patient

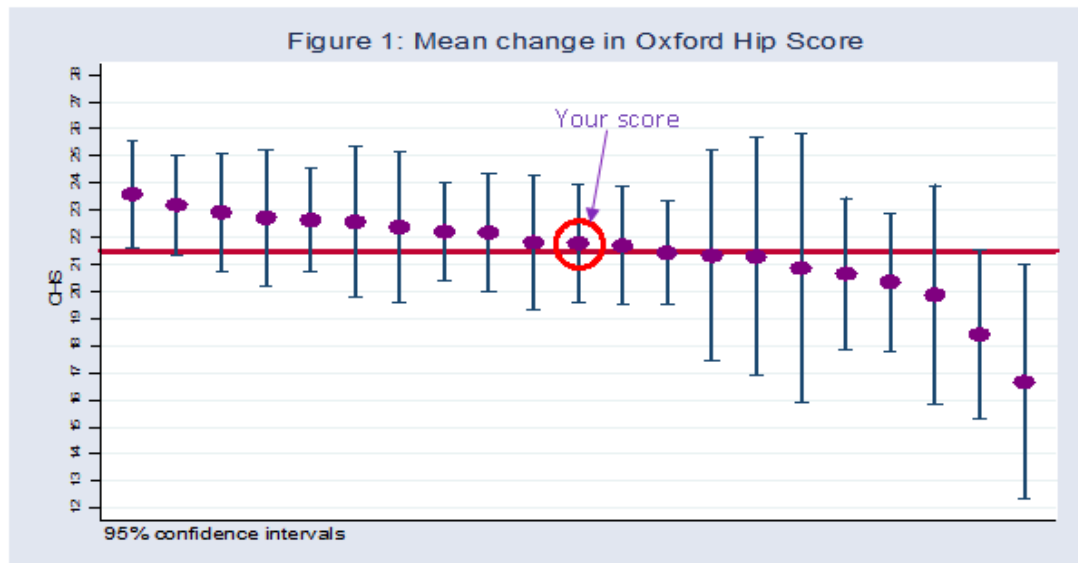
Reported Outcome: Feedback Interpretation and Learning Experiment (PROFILE) trial—refer to (ISRCTN 69032522) for more details. PROFILE trial aimed to evaluate the effectiveness of the NHS PROMs Programme methodology for surgeon level feedback in an Irish context (1). In brief, PROFILE tests the hypothesis that healthcare professionals who receive benchmarked PROMs feedback will have better future outcomes than those who do not receive feedback. This trial was undertaken in Ireland where performance monitoring has not yet progressed beyond measuring processes such as waiting times, length of stay, and adherence to hygiene standards. This was the first time the participating surgeons had received peer-benchmarked feedback about their patient outcomes. We discuss the methodology of the trial below and describe the nature of the PROMs feedback provided to clinicians within the trial. We then subsequently explain the methodology of the nested qualitative study.

PROFILE is a trial of 21 high-volume hip replacement surgeons and their patients. In the trial, patients were asked to fill out a questionnaire before and six months after their operation. Questionnaires included demographic questions on the patient's age, gender, duration of symptoms, and the PROMs included were the Oxford Hip Score (OHS) (231), the EQ-5D (251), a shortened version of the Hip Osteoarthritis and Outcome Score (HOOS) (252), and a general health status item. Post-operative questionnaires were similar except they also included questions on the results of the operation and post-operative problems, including allergy or reaction to a drug, urinary problems, bleeding, and wound problems (1). Pre-operative data collection took place in a pre-assessment clinic, if available, or alternatively when the patient was admitted to the hospital for surgery. The data collectors included nurses and registrars. Post-operative data collection was managed by the research team using a

postal survey. Questionnaires were posted to patients six-months after their surgery, and a reminder was sent four weeks later if a reply was not received within this timeframe.

The data collection occurred in two phases: pre- and post-feedback. The pre-feedback phase was used to generate peer benchmarked PROMs reports for the 11 surgeons randomised to the intervention arm of the trial. The content of the feedback report was based on research which examined clinician's preferences on metrics used to compare surgical performance (230) and included the mean change (post-operative minus pre-operative) in the OHS, the proportion of patients that reported improvements in their hip problem, and the proportion of patients that reported having at least one of four problems after surgery. Case-mix adjustment of the OHS was used to ensure a fair comparison of surgeon level results. The OHS was adjusted to account for patients' pre-operative OHS, age, gender, general health status, and mental health status. Surgeon's scores were clearly highlighted for each outcome demonstrating how they performed in comparison to the other 20 surgeons in the trial; however the identity of these surgeons remained anonymous (Figure 8). The feedback report was based on data from 573 patients. A minimum patient recruitment was set at 32 patients per surgeon—a requirement that was necessary to accurately benchmark outcomes. The post-feedback phase of the trial follows the same data collection procedures on a new cohort of patients. In this phase, PROMs act as the outcome measure by examining differences between the feedback and control arms. Follow-up data collection for PROFILE is currently ongoing, and the results will be published in late 2014. Feedback was provided in January 2013 and the interviews were performed between three and five months later.

Figure 8: Example of peer benchmarked PROMs feedback



The qualitative study

This paper reports on a qualitative study that was nested within the PROFILE trial described above. The qualitative study employed a Framework approach (253). This is appropriate when aiming to generate policy-orientated findings and recommendations for practice in a field where an existing conceptual framework derived from the literature was an appropriate starting point for the data collection and analysis (254-255).

Sampling and data collection

All 11 surgeons in the feedback arm of the PROFILE trial were invited to participate in a face-to-face interview, and consented to do so. Given that this represents a complete capture of all possible respondents of interest, the sampling method can be characterised as a census. The participants varied in terms of the setting of their usual workplace, their relative performance ranking and their previous experience of using PROMs (Table 9). The 10 surgeons in the control arm were not interviewed because they did not receive feedback, so their reactions to this information could not be elicited.

The topic guide was informed by the objectives of the research and the results of a systematic review undertaken by the authors to synthesize existing qualitative evidence on professional's experiences of using PROMs as quality improvement tools (142). This review identified four themes: practical considerations, attitudes towards PROMs, methodological concerns, and the impact of the feedback on care. A draft discussion guide was developed from these themes. This was reviewed by the research team and independently with clinical professionals before finalising the discussion points. The final guide covered five topics: experiences of using PROMs, attitudes towards using PROMs as peer benchmarking tools, methodological factors, practical factors with collecting and using PROMs data, and the impact of the information on behaviour (Appendix 14).

The interviews were performed by MB, who is a trained health services researcher with seven years' experience working in both qualitative and quantitative approaches that reflects a pragmatic paradigm underlying this research. Before commencing each interview, the rationale for the study and the specific purpose of the discussion was clearly outlined to participants. Each surgeon provided written consent for digital recording and verbatim transcription. The study was conducted according to ethical guidelines (256). The Research Ethics Committee of the Cork Teaching Hospitals (CREC) approved the study protocol, as well as the ethics committees within the hospitals.

Data analysis

A Framework approach was employed to analyse the data (253). Framework analysis uses a stepwise approach to ensure a systematic, rigorous, and transparent approach to the analysis (255). QSR International's NVivo 10 software was used to assist with the analysis (197). First, the raw data were repeatedly read to identify initial

concepts. A 'one sheet of paper' mapping exercise (257) developed these ideas into a preliminary framework. This framework was tested by labelling (indexing) a sample of the data, and was revised before being populated by the entire dataset. Next, the data were categorised and synthesised by sorting and summarising the material into charts. The raw data were exported into these charts to ensure the meaning and context of the participants views were retained. Lastly, patterns within the data were examined to help describe and explain the findings by sequentially comparing each theme against the other four themes and across different cases (253). The typology emerged from two themes (attitudinal and impact), and differences between the remaining three themes (conceptual, methodological and practical) were examined against the typology. A framework was developed to describe the relationship between the themes by examining subtle differences across the three types of participants. The characteristics of the participants were examined in a similar manner to produce explanations for the groupings.

An academic clinician independently coded three randomly selected transcripts and helped develop and refine the framework prior to commencing the indexing. As the authors are not clinicians, this perspective ensured the analytic framework evolved with a sensitivity to the culture of the Irish healthcare system. JB and JG participated in discussions about the analytic framework throughout the process. Regular analysis meetings between the authors challenged the analytic process, interpretation of the data, and any possible observer bias. MB kept a reflective journal during the analysis and used personal memos to track decisions and challenge any personal or professional biases in interpreting the data.

Given the influential position of surgeons within the healthcare service, it has been noted that relatively few participants (between six and twelve) may offer deep

insights into the structure and culture of the system (258-259). Therefore, the framework was developed after eight interviews were completed. The final three interviews were used to examine saturation. This was undertaken by comparing the themes emerging from each additional interview against the framework to establish if any new issues or concepts emerged (260).

Rigour

We took a number of steps to enhance the trustworthiness of the study finding (261). First, we examined previous research to frame the findings. Second, we built trust with the study participants by clearly explaining the research aims, declaring the researcher's independent affiliation to the HSE or governing bodies and assuring the interviewees that their confidentiality would be maintained. Third, we sought peer scrutiny throughout the study by involving healthcare professionals when drafting the discussion guide, checking shared meaning of concepts by jointly coding transcripts with an independent clinician, and sharing ideas with the research team throughout the development of the framework and when defining themes. Fourth, the lead interviewer maintained a reflective approach throughout the research by writing a journal during the data collection phase and keeping memos throughout the analysis phase. Transferability was enhanced by recruiting participants from 16 organisations across mixed settings and with mixed levels of experience, and by providing rich information on the study context and findings to enable future researchers to draw comparisons. Dependability was enhanced by clearly describing our methods to enable study replication. Confirmability was promoted by recognising study limitations and by declaring the researcher's beliefs and assumptions (261).

5.4 Results

All 11 consultants in the feedback arm of the trial agreed to participate. All participants were male, six worked in a public setting, five in both public and private settings, and one in a private setting only. Six surgeons had above average OHS scores and five had below average OHS scores when all surgeons were benchmarked against each other. Two had moderate experience, six had minimal experience and three had no experience of previously using PROMs (Table 9). Interviews were held privately in the participant's workplace. The median length of the interviews was 42 minutes (range 15 to 84); the longer interviews tended to focus more thoroughly on the methods.

Table 9: Characteristics of participants

Sex	Setting	Above/Below average (OHS)	Experience of using PROMs
Male	Public	Above	None
Male	Public	Above	Minimal
Male	Public	Below	Moderate
Male	Public	Below	Minimal
Male	Public	Below	Minimal
Male	Private	Above	Minimal
Male	Mixed	Above	Moderate
Male	Mixed	Above	Minimal
Male	Mixed	Above	None
Male	Mixed	Below	None
Male	Mixed	Below	Minimal

Five themes were initially identified: conceptual (understanding PROMs), methodological (focus, accuracy and interpretation of the data), practical (issues with

collecting and using the data), attitudinal (valuing the information), and impact (using the information to make changes to the processes of care). Subsequently the themes about ‘attitudes’ and ‘impact’ were merged due to their co-dependency on participant’s reactions to the feedback. Quotations were selected to represent the essence of each sub-theme and have been coded to protect the subject’s confidentiality (Table 10).

Table 10: Themes, sub-themes and excerpts from the participants

Themes	Sub-themes	Excerpts
Conceptual	Subjective measurement	<p>‘Getting patients to fill out forms is grossly inaccurate in my book...the patient 9 time out of 10 wouldn’t understand what hip pain is’ (S9)</p> <p>‘There is some subjective element but it is a reasonably validated objective assessment’ (S2). ‘Well they are partly objectified, aren’t they?’ (S11)</p> <p>‘I suppose the difference maybe with my results is the difference between the maybe more objective measures and the subjective measures’ (S5)</p>
	PROMs V Satisfaction	<p>‘Patient satisfaction in a sense is a balance between what their expectations were beforehand and what they achieved afterwards’ (S10)</p> <p>‘You know there is one outcome there on how much the patient likes the outcome as I like to call it’ (S2)</p> <p>‘When they are not perfect, they manifest that by saying they are quite poor’ (S7)</p>
	PROMs V clinical data	<p>‘Clinically I see very very very few problems and very few dissatisfied patients...that is just wrong. I am sorry I just can’t accept that’ (S10)</p>
Methodological	Focus and variability	<p>‘You should concentrate on operations that have dubious results’ (S8)</p> <p>‘The increments between each surgeon are tiny ...I mean your spread there between top and bottom is only six points’ (S7)</p>
	Timing	<p>‘To see if there was any differences at four to six weeks’ (S4)</p> <p>‘The other thing is the timing is critical because one would generally not measure anything in hip surgery and knee surgery for at least one year’ (S11)</p>
	Choice of measures	<p>‘That score has issues with validity for certain age groups’ (S1)</p> <p>‘The patient might perceive it as a complication but it is not, it is part of the normal process’ (S8)</p> <p>‘You know it has to be patients with a problem after surgery that is directly related with the surgery’ (S10)</p>
	Interpretation	<p>‘Unless I was able to compare myself against somebody else who does things quite differently’ (S2)</p> <p>‘I mean strictly speaking someone that is at the tail end should be at the tail end in all three’ (S7)</p>

Themes	Sub-themes	Excerpts
	Validity (data quality, case-mix adjustment, sampling)	‘Something is wrong somewhere: either they have problems and they are not telling me or else there is something odd in data collection’ (S10) ‘Even if you adjust them it is not going to give you the proper information’ (S1)
Practical	Time	‘If I had time, maybe. I don’t have time. I mean, I have continuous ideas...and am...let’s say resolutions to measure outcomes better and more often and all the rest of it but we don’t have the time like and we don’t have the staff’ (S11)
	Support	‘No interest. No support. No help. No funding’ (S2) ‘We don’t have anything strictly audit related because the big problem with the hospital audits is the information gathering is poor’ (S7) ‘You need generally a political will to get it because it can achieve nothing but to cost them more’ (S2) ‘You need software, you need somebody to analyse it’ (S3) ‘...that takes help, statistical help’ (S4)
Attitudinal	Value	‘There have been a lot of high profile problems in recent times and maybe these kind of problems would have been spotted sooner if we were collecting this type of data’ (S5) ‘You see your patients and they are happy but in general terms you don’t know how you are performing compared to your peers’ (S4)
	Undecided	‘That is kind of a relatively disappointing figure, I would have thought and not just mine, I think the overall is kind of a little bit disappointing. Why it is? I am not sure’ (S3)
	No value	‘I just think there is a lot of effort being put in there for not a lot of surgical gain from my perspective’ (S8).
Impact	Impact	‘I am going to try and do it better’ (S4) ‘I went off for a few days and started thinking about things so even though my results would appear not to be brilliant, it was very beneficial for me’ (S7)
	No impact	‘I seem to be in the middle there and I wouldn’t be changing what I do on the basis of it’(S2) ‘Unfortunately, it does not provide me with one iota that helps me make my next score any better’ (S10)

Theme one: Conceptual—understanding PROMs

Participants varied in their understanding of PROMs as a concept. This became evident in three ways: comprehending subjective measurement, confusing PROMs with patient satisfaction measures, and aligning PROMs with clinical data.

Subjective measurement

Participants declared a respect for eliciting information from patients, but expressed concern about the scientific properties of PROMs. There was an underlying doubt about patient's ability to report on issues such as pain and physical function.

Surgeons consciously deliberated the concepts of subjectivity and objectivity in relation to the constructs being measured. PROMs data were seen by many as 'subjective' constructs and therefore less trustworthy. However, the distinction was not absolute as they ranked different PROMs by their level of 'objectivity.'

PROMs versus satisfaction

Consultants often did not distinguish the difference between PROMs and measures of patient satisfaction or experience, and thus assumed that the questionnaires captured information on the processes of care throughout their healthcare journey.

PROMs versus clinical data

Participants expected PROMs data to align closely with clinical indicators. Many expressed disbelief about the percentage of patients who reported that they had not improved or had a problem after surgery. Surgeons felt that these figures did not match their experience of clinical practice and verbal feedback from patients post-operatively.

Theme two: Methodological—measurement decisions, measurement accuracy and interpretation

This theme captured methodological issues around the focus of measurement, the timing of data collection, the choice of measures, trust in the accuracy of the PROMs feedback, and problems with data interpretation. A key underlying issue within this theme were the methodological threats to the trustworthiness of PROMs as an indicator of surgeon's performance.

Focus of measurement

Participants questioned the rationale for focusing on hip replacement surgery. One consultant queried the cost-effectiveness of concentrating on a procedure where poor outcomes are perceived to be rare. Some surgeons discussed the relatively small variability between surgeons and, therefore, the clinical value of performance management in a field where only marginal improvements may be possible at the population level.

Timing

Participants discussed the timing of the post-operative data collection. One participant was interested in the rate at which patients recover from different approaches and techniques, and how this would influence performance ranking at different time points, particularly in the short term. Others believed that six-month follow-up was too soon because patients continue to improve for up to a year.

Choice of measures

Participants also recognised that the measures collected influenced the value of feedback. One surgeon questioned the appropriateness of the OHS because it was developed for an older population with arthritic problems. The choice of measures

became particularly pertinent when participants considered the data about post-operative complications. Some felt that it was unfair to associate these complications with their performance because they believed that the specific problems in question were not a direct complication of surgery.

Accuracy of the feedback

Interviewees expressed a range of opinions about the validity of PROMs. The factors identified were related to possible biases, confounding, and chance.

Participants were aware that incorrect administration and completion of the measures would affect the data quality. In particular, they were concerned about the potential to manipulate scores by failing to recruit patients who may be more likely to have a poor outcome, thus creating a selection bias. Incorrect completion of the measures was identified as a possible source of information and recall bias. Participants questioned the patient's ability to complete the PROMs correctly. This was considered especially relevant for patients with co-morbidities who might confuse problems arising from their hip osteoarthritis with problems arising from other conditions. Concern was also expressed about the possibility that patients with low literacy might tick random answers or ask family/friends to complete the questionnaire on their behalf. Participants were also worried about the influence of patient expectations on PROMs, which might lead to an underestimation of the 'true' outcome. Others argued that patients might deliberately underestimate their pre-operative outcomes in the belief that the information was being used to ration care. However, one participant identified a scenario where patients may overestimate their outcome due to a 'post-event rationalisation,' where patients start to justify their choice to have the operation, resulting in a belief that their outcome is better than it actually is.

The issue of confounding was identified as a serious threat to the accuracy of the findings. Consultants were concerned about the impact that patient case-mix, differences in resources across hospitals and differences in support services at a community level had on patient outcomes. Patient level confounding was perceived as the most serious threat and many were sceptical about the accuracy of adjusting for case-mix.

Lastly, some surgeons were concerned about the influence of chance on findings. Surgeons were interested to see if their ranking would be similar with a larger sample or different samples of patients. Therefore, many were keen to receive additional feedback reports to monitor their performance.

Interpretation of the feedback

Consultants had difficulty making sense of the PROMs feedback. Understanding the variation between and within surgeons was challenging. Surgeons also found it hard to identify opportunities for quality improvement within the feedback.

Consultants had problems identifying reasons for variation between surgeons because of the number of causal factors linked to PROMs. Participants found that the PROMs feedback alone was insufficient to provide explanations for poor performance. However, some thought that linking PROMs to information about clinical practices might improve future decision making. Finally, some aspects of the feedback confused certain participants who ranked differently across the outcome measures because they could not explain the reasons for such deviations.

Theme three: Practical issues with collecting and using the data

The process of collecting and using PROMs data created barriers to a positive engagement with the exercise. Data collection added to workload pressures. Many

surgeons stated that their support staff were not willing to accept the increased workload associated with questionnaire administration. Furthermore, surgeons recognised that political will at a hospital and system level was necessary to maintain such initiatives because real quality improvements often require a level of resource flexibility. In addition, there was concern that both clinical and managerial professionals lack the knowledge and training to use PROMs data. Surgeons recognised that in the absence of such training there was a danger that the data may be inappropriately used.

Typology: attitudes (valuing the data) and impact (using the data)

Three distinct groups emerged with respect to views about the final themes: attitudes (the value attached to PROMs) and impact (the likelihood of using PROMs to change clinical practices). Two surgeons (Advocates) expressed a positive attitude to the feedback they received and stated that the information had an impact by promoting a reflective process focusing on their clinical practice, although they did not explicitly state specific changes to the process of care. One of these surgeons stated that the results provided additional motivation to continuously aim to perfect his technique. The other stated that the results promoted a process whereby he considered at depth the aspects of care that may have affected performance.

A separate group of four surgeons (Converts) were uncertain about the value of PROMs, and this reduced their inclination to use the data. They lacked the knowledge to make an informed decision on the usefulness of PROMs but were reassured that their performance was similar to their peers. This group generally felt that it is important to know what patients think about their outcome but emphasised the need to provide actionable feedback to professionals.

A third group of five surgeons (Sceptics) believed that the PROMs feedback they had received was not clinically useful and so the feedback had no impact on their behaviour. They felt that there were too many scientific concerns to trust the data, that the data collection was cost-ineffective, and that the data were not a useful source of ideas about ways to stimulate improvement.

Relationship between themes—a conceptual model

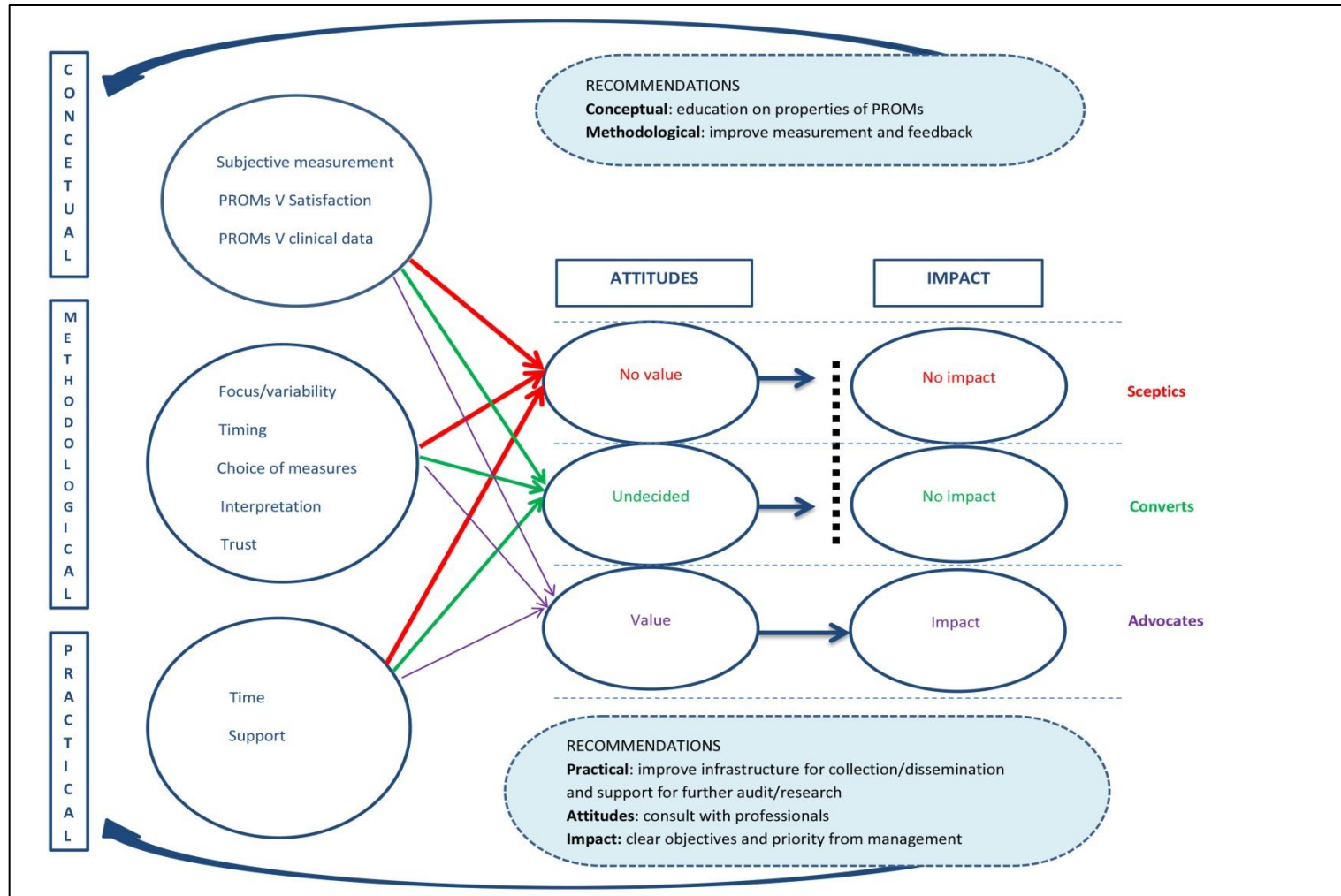
A matrix helped examine patterns in the themes (Table 11). By examining the patterns between the themes and the typology, it became clear that the conceptual, methodological, and practical issues were important determinants of professional's attitudes towards PROMs. The attitudes, in turn, defined the impact of the information on behaviour. A conceptual framework was developed to depict the relationship between the themes (Figure 9).

Table 11: Mapping of themes and sub-themes across surgeons

Surgeons	Characteristics	Typology	Conceptual	Methodological	Practical
Surgeon 4	Baseline performance (OHS): Above average Setting: Mixed Experience: Moderate	Advocate (value and impact)	PROMs V Satisfaction	Interpretation Timing Validity	Support/infrastructure
Surgeon 7	Baseline performance (OHS): Below average Setting: Public Experience: Minimal	Advocate (value and impact)	PROMs V Clinical	Interpretation Focus/variability Validity	Support/infrastructure
Surgeon 2	Baseline performance (OHS): Below average Setting: Public Experience: Moderate	Convert (undecided and no impact)	Subjective measurement PROMs V Satisfaction	Interpretation Validity	Time/workload Support/infrastructure
Surgeon 3	Baseline performance (OHS): Above average Setting: Private Experience: Minimal	Convert (undecided and no impact)	PROMs V Satisfaction PROMs V Clinical	Interpretation Focus/variability Validity	Time/workload Support/infrastructure
Surgeon 5	Baseline performance (OHS): Above average Setting: Public Experience: None	Convert (undecided and no impact)	Subjective measurement PROMs V Satisfaction	Interpretation Focus/variability Timing Validity	Time/workload Support/infrastructure
Surgeon 6	Baseline performance (OHS): Below average Setting: Public Experience: Minimal	Convert (undecided and no impact)	Subjective measurement	Interpretation Validity	Support/infrastructure
Surgeon 1	Baseline performance (OHS): Above average Setting: Public Experience: Minimal	Sceptic (no value and no impact)	n/a	Interpretation Measurement Timing Validity	Time/workload Support/infrastructure

Surgeons	Characteristics	Typology	Conceptual	Methodological	Practical
Surgeon 8	Baseline performance (OHS): Below average Setting: Mixed Experience: None	Sceptic (no value and no impact)	Subjective measurement PROMs V Satisfaction	Interpretation Focus/variability Validity	Time/workload
Surgeon 9	Baseline performance (OHS): Above average Setting: Mixed Experience: None	Sceptic (no value and no impact)	Subjective measurement	Interpretation Focus/variability Timing Validity	n/a
Surgeon 10	Baseline performance (OHS): Below average Setting: Mixed Experience: Minimal	Sceptic (no value and no impact)	Subjective measurement PROMs V Satisfaction PROMs V Clinical	Interpretation Focus/variability Timing Validity	Support/infrastructure
Surgeon 11	Baseline performance (OHS): Above average Setting: Mixed Experience: Minimal	Sceptic (no value and no impact)	Subjective measurement PROMs V Satisfaction PROMs V Clinical	Interpretation Focus/variability Timing Validity	Time/workload

Figure 9: Conceptual framework of the relationship between themes



There was evidence that surgeons' understanding of PROMs was an important determinant of the extent to which they might value and use the data. The 'Converts' and 'Sceptics' were more likely to deliberate the distinction between subjective and objective measurement, placing more trust in the scores that were perceived to be more 'objective,' and were more likely to misinterpret the information.

The strongest influence on surgeon attitudes and behaviour was the methodological theme. The 'Advocates' focused less on the factors that may impact on the data quality and more on further research opportunities to investigate the reasons for variations in outcomes, such as examining the relationship between outcomes and expectations, exploring rankings at different time periods, and undertaking case-study reviews. The 'Converts' tended to appreciate aspects of the feedback but were perturbed by some of the methodological issues. Their discussion focused in more detail on the possible errors in the data, particularly the impact of incorrect administration and completion of the questionnaire on data quality. Similarly, these professionals highlighted inconsistencies between the PROMs scores deliberating whether the divergences were associated with inaccuracies in the data. The 'Sceptics' focused on reasons why they did not trust the data. They also highlighted the impact of incorrect completion and administration on findings, and questioned the measurement properties of PROMs, the focus on hip replacement surgery, and the complexity of causal factors determining outcomes.

The views of the groups also differed with respect to their concerns about practical issues. The 'Advocates' focused on how PROMs could be used more effectively if there was greater audit and research support. The 'Converts' focused on the impact on workload, the lack of collaboration between staff and management, and the cost of data collection. The 'Sceptics' provided an insight into the negative consequences

of collecting PROMs, including the opportunity costs involved, and were cynical about the willingness and ability of their local hospital to support real quality improvements.

There was no obvious relationship between surgeon responses and their performance ranking or the setting in which they worked. However, previous experience with using PROMs may have influenced their responses. Two surgeons had experience of collecting PROMs routinely in practice: one of these was classified as an ‘Advocate’ and one as a ‘Convert.’ Five surgeons had minimal experience of using PROMs for research purposes: one was classified as an ‘Advocate,’ two as ‘Converts’ and two as ‘Sceptics.’ Three surgeons claimed they had no experience with using PROMs: one was classified as a ‘Convert’ and two as ‘Sceptics.’

5.5 Discussion

This is the first study of healthcare professionals’ experiences of receiving peer-benchmarked feedback using PROMs. Three groups of surgeons emerged from the analysis: Advocates, Converts, and Sceptics. ‘Advocates’ had positive attitudes towards the use of PROMs and admitted that the information had an impact on their behaviour by promoting a reflective process on their clinical practice. ‘Converts’ had mixed attitudes because they were uncertain about the value of PROMs, which prevented them from using the data to inform their practice. ‘Sceptics’ portrayed negative attitudes towards the value of PROMs and reported that the feedback had no impact on their behaviour. The barriers towards the use of PROMs information may be categorised into conceptual, methodological and practical factors.

Conceptual issues refer to problems with understanding PROMs, for example, comprehending subjective measurement, confusing PROMs with patient satisfaction

measures, and aligning PROMs with clinical data. These problems were more common among the ‘Converts’ and ‘Sceptics,’ which may be partly linked to an unfamiliarity with using these measures. Though based upon a small sample size, this is tentative evidence that familiarity with PROMs is associated with a more positive disposition towards their use. Methodological concerns, for example, the focus of measurement, the timing of data collection, the choice of measures, the validity of the information, and interpretation of the data were further barriers to full engagement with PROMs. The ‘Advocates’ used the information to prompt ideas for further investigations. In contrast, ‘Converts’ and ‘Sceptics’ were more likely to question the data quality and less likely to accept responsibility to further explore the reasons for variations in performance. Finally, practical constraints such as workload pressures and a lack of support were also barriers towards the uptake of PROMs. Practical issues were more of a concern for the ‘Converts’ and ‘Sceptics.’ This may be because the ‘Advocates’ already had some of these processes in place. However, implementing the routine use of PROMs not only requires dedicated staff time for data collection but also appropriate information technologies, statistical support, and resource flexibility to appropriately use the information, which can be difficult to procure.

Implications of findings

These findings outline the barriers to the effective implementation and use of PROMs in practice. The conceptual framework produced by this research can be used by practitioners, managers, and policy makers who hope to use PROMs benchmarking to improve the quality of care and by researchers who are interested in the implementation of these strategies.

Some participants were familiar with using PROMs for research projects or had experience collecting PROMs in practice to manage patient care; however the use of PROMs as performance measures was a new concept for most of the surgeons. This inexperience may have led them to make sense of PROMs by relating or equating them to measures they were familiar with in a performance monitoring context, such as clinical indicators like revision rates and patient satisfaction surveys. However, these were not measured in this study. PROMs address unique constructs and perform a unique role in health measurement (262-267). These findings highlight that providing training on the different functions of PROMs, the measurement properties of the instruments and the interpretation of the data is necessary if PROMs are to be effectively used in practice. Furthermore, co-designing feedback reports with professionals would generate information that professionals perceive as useful and increase the likelihood of positive engagement (268-269). Further qualitative research could be used to assess whether opinions of surgeons change as they receive PROMs feedback and become more familiar with the data.

The research highlights many interesting methodological questions for future research studies. First, the recent application of PROMs as performance monitoring tools creates uncertainty regarding the adequacy of the existing measures. Many of the tools were developed to assess the effectiveness of healthcare interventions across patient populations, but have been subsequently applied in clinical practice for individual patient-level evaluations and to detect differences in quality of care between healthcare professionals (249). This creates problems as the reliability and validity of the information generated for these different uses cannot be guaranteed (183, 238).

A second issue to consider is that PROMs data are not directly ‘actionable’ in that they do not point to solutions that will improve the quality of care. PROMs produce scale-level data that summarise the responses to a number of items. Scale-level data, although improves reliability by asking a number of questions in relation to a particular construct, can be more difficult to interpret from a clinical perspective and are more suited to establishing ‘that’ differences exist as opposed to ‘why’ they exist. A possible solution to these measurement and interpretation issues is to adopt psychometric techniques such as Rasch modelling. Rasch analysis has the capability of producing more precise measurement instruments and reliably enables the interpretation of the information at the item and scale level by linking each item to a score on the scale. This can help clinicians decipher the implications of a change in score. For example, if each item on a scale is linked to a score on a scale which measures function, a change score is no longer an abstract number. Instead, clinicians can link scores to the location of items on the scale. This may inform clinicians that a patient’s maximum function changed from having difficulty standing up from a sitting position to being able to climb a stairs without difficulty. Clinically this information is more intuitive than summary scores (223, 270).

Third, surgeons identified the need to produce meaningful and useful feedback suggesting that PROMs data should be provided alongside clinical and patient experience data. This information may offer an insight into the factors causing variation. Our knowledge about how these perspectives correlate across the range of measures is not well advanced. For example, a review examining the relationship between satisfaction with care and PROMs found a positive correlation, however the causative direction of this relationship could not be determined (271). The evidence on the relationship between improving processes of care and outcomes is also weak

(272). This may be a symptom of inadequate efforts to generate high data quality and to test the use of these measures in practice prior to their routine introduction.

However, it is important to recognise that ongoing developments in both process and outcome measures and measurements are necessary to drive a deeper conceptual understanding of the link between these elements of care (272-273).

Fourth, the focus of measurement also needs to be considered, as performance monitoring will have the greatest impact when the variation between professionals is large or baseline performance is poor. Hip replacement may not be the most sensible procedure to target, as this study found that the variation between surgeons was small and baseline performance was good (239).

Fifth, the wider outcomes literature has identified some additional attributes of successful performance improvement initiatives (274-275). There is evidence that a meticulous focus on generating high-quality data can promote positive changes in outcomes over time, particularly for 'bad outliers' (274), and that collaborative improvement programmes can stimulate improvements far more quickly than efforts by single providers (275). The benefit of a collaborative programme is that large sample sizes enable a robust assessment of relationships between process and outcomes, identifying best practices that can be rapidly rolled out to the entire group. This in combination with an increased focus on creating an appropriate environment for quality improvement can lead to better patient outcomes (275). Our study similarly highlights that building for a momentum for change depends on effective leadership and ongoing practical support to help professionals identify where improvements are required (226).

Study limitations

There are some limitations to this research. First, the research is based on the views of only eleven participants. It should be acknowledged that consultants are an ‘elite’ source of insight, given their authority and in-depth knowledge of the system (259). In addition, established methods were used to assess if data saturation was reached (260). Nevertheless, the generalisability of the findings to other types of healthcare professionals should be considered and further testing of the typology with a larger number of participants is advisable to advance our understanding of the reasons for different reactions to the data. Also, it may be possible that one may find additional categories to the typology if explored in a larger number of participants or on varied types of professionals. Second, the impact of performance measurement is dependent on various contextual factors such as local culture and governance structures. This research was undertaken in Ireland, where professional performance assessment is still at a rudimentary level; therefore professionals may have had a general suspicion of peer benchmarking. Third, the research is based on only one round of feedback. Professionals may be more likely to engage with PROMs data if they receive regular feedback reports and can observe meaningful trends over time. Fourth, qualitative research will not capture the psychological impact of measurement on behaviour such as the Hawthorne effect, which may lead to more subtle changes to practice. Finally, this research does not explore the influence of feedback on the wider healthcare system. The NHS PROMs programme provides feedback at the NHS Trust level that engages different aspects of the clinical governance infrastructure and may provide useful information to different actors such as patients and purchasers.

5.6 Conclusion

Interest in the use of PROMs as quality improvement tools is growing. However, this research demonstrates that there are conceptual, practical, and methodological issues that determine attitudes towards the use of PROMs and, in turn, professionals' willingness to use the information to inform practice. Policy makers and researchers need to engage more effectively with professionals, provide sufficient education and training, develop better measures and feedback mechanisms, and help to build a more supportive and efficient data collection infrastructure.

6

Discussion

Chapter 6- Discussion

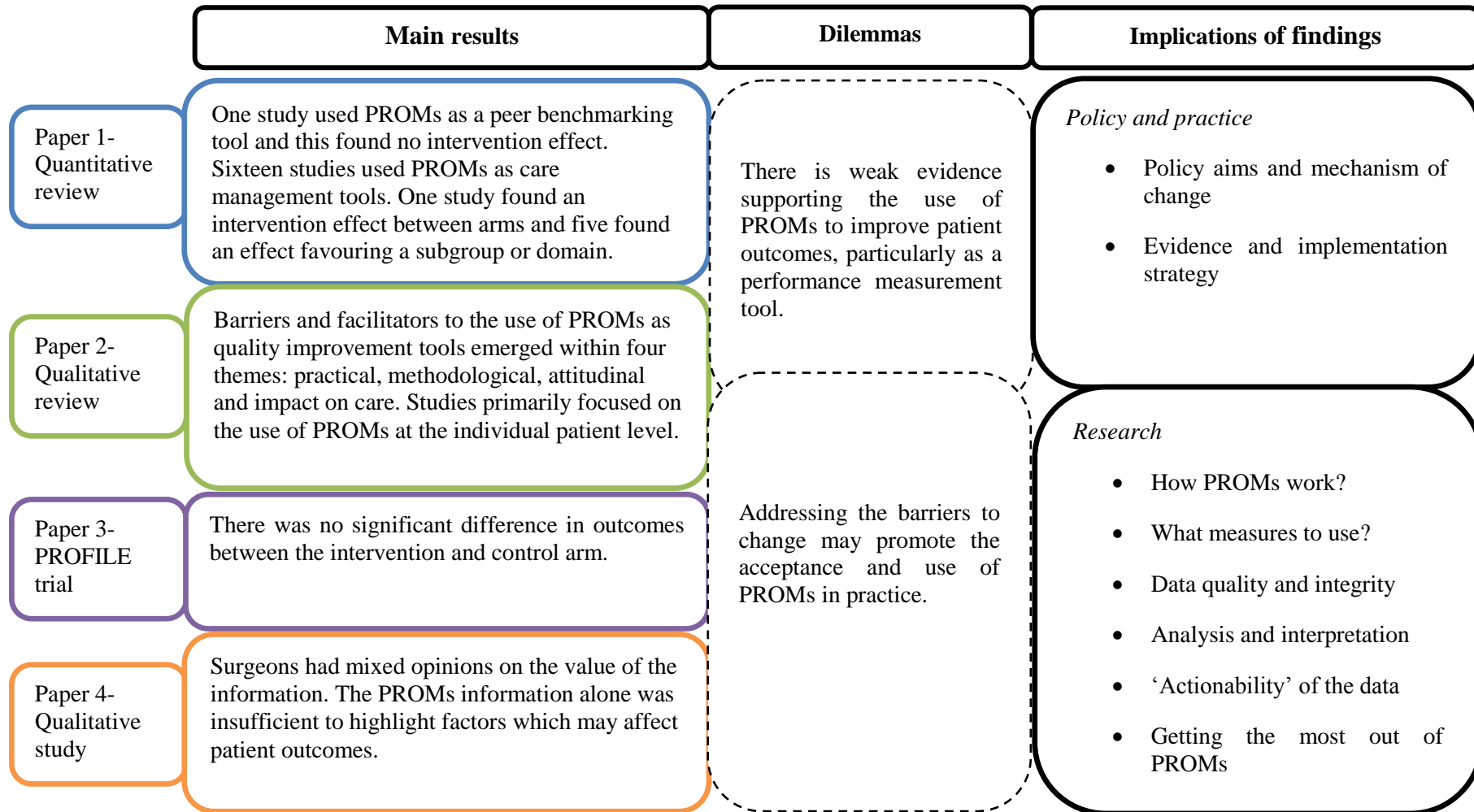
The use of PROMs as an audit and feedback intervention is a relatively novel quality improvement approach. This research adds significantly to knowledge in this field. Prior to this research, there had been no systematic review that examined the impact of feedback given the different functions of PROMs (as an individual patient management tool and a performance measurement tool); there had been no systematic review that synthesised the qualitative evidence on professionals' experiences of using PROMs information; there had been no randomised controlled trial that examined the impact of providing peer benchmarked PROMs feedback to surgeons and there had been no qualitative research that explored surgeons' reactions to this feedback.

This discussion summarises the findings from the four research papers (Chapters 2-5) presented in this thesis and outlines how this new evidence will inform future considerations for the implementation of this quality improvement intervention. This is followed by a discussion of the limitations of the research before the main conclusions are presented.

6.1 Summary of main findings

In this section, the findings of each paper will be summarised (Figure 10) while drawing links between the studies to highlight the implications of this research for future policy, practice and research.

Figure 10: Overview of results, dilemmas and implications of findings



Quantitative systematic review

In Chapter 2, the published evidence on the impact of feeding back PROMs information to providers was examined. This systematic review differed from previous reviews as it specifically investigated the value of PROMs feedback about individual patients (as a patient management tool) and about groups of patients (as a performance management tool), and examined the effect of feedback on patient-reported health outcomes. The review found weak evidence to support the hypothesis that providing PROMs feedback improves patient outcomes. Of particular importance, our review found that the evidence on the use of PROMs as a performance measurement tool is scant as only one relevant study was identified and this found no intervention effect. In fact, these results suggested that the functional status deteriorated for all participants over the course of the study. This study included 48 primary care physicians in California. All physicians were informed that their elderly patients would be monitored but only the physicians in the intervention group received aggregated peer-comparison feedback. Functional status was assessed using the Dartmouth Primary Care Cooperative Information Project chart method and data was collected over two time periods, the first in 1992 and the second in 1995. The limitations of this study were that it included physicians from one group-model health organisation in America, it provided only one round of feedback to physicians, and the length of time between the two data collection periods was considerable which may explain the decline in function over a three year period for the elderly cohort (179).

Sixteen studies were identified which focused on the use of PROMs as a patient management tool. Only one of the studies found a positive intervention effect (175) and an additional six studies found positive results favouring the intervention group

for a particular subgroup or domain (163, 169, 172-174, 178). The quality of studies was generally poor, and those studies which found an intervention effect were of slightly lower quality than those that did not find an effect. This raises concerns over the strength of the findings. There is tentative evidence that PROMs feedback may be more valuable when used as a management tool for patients with specific healthcare problems. This may be due to the ‘actionability’ of the information for certain conditions (the ability to identify a specific course of action) or it may be linked to a greater potential for improvement given the severity of the problem. Since this review was published, no further publications that meet the review’s inclusion criteria have been identified. However, two reviews were published in 2013 that explored the impact of routinely collecting PROMs in an oncology setting (189, 276). These reviews also concluded that there is weak evidence to suggest that PROMs feedback improves patient outcomes.

Qualitative systematic review

In Chapter 3, the evidence on professionals’ experiences of using information from PROMs feedback was examined. For the first time, this review detailed the barriers and facilitators to the use of PROMs in practice. Thus, the findings provide unique insights into the challenges associated with implementing and using PROMs from the healthcare professionals’ perspective.

The review identified sixteen qualitative studies. Similar to the quantitative systematic review, studies primarily used PROMs as a patient management tool. However, two of the studies provided feedback at both the individual and group level, thereby using PROMs as a patient management tool and a performance measurement tool. Both papers are based on the introduction of routine outcome measures in the mental health service in Australia (277-278). Callaly et. al employed

focus groups (n=13) and interviews (n=7) to explore the attitudes of nurses (n=64), allied health professionals (n=12) and medical staff (n=7) to the implementation and use of outcome measures two years after their routine introduction into care. Patients completed the BASIS-32 on a touch screen computer and the results were immediately available for professionals to monitor progress and to assist in managing care. The data was also aggregated to the level of the provider and publically reported to drive quality improvements. The study found a mixed level of acceptance in relation the validity and usefulness of the measures. However, patient reported measures were perceived to be more valuable than clinician reported measures. Professionals believed that systems to support the use of outcome measures were required, as well as ongoing training (277). Meehan et. al employed the use of focus groups (n=34) to explore reactions of mental health staff (n=324) to the introduction and utility of outcome measures in clinical practice eight months after implementation. Patients completed the Mental Health Inventory on a computer which enabled the automatic generation of patient level feedback or summary reports for benchmarking purposes. Once again mixed views were reported on the perceived value of the measures in practice, but in particular many expressed ambivalence towards the measures. The findings highlighted barriers towards the use of PROMs which included competing work demands, lack of support from senior staff and fear of how the data might be used by management. These factors prevented staff from fully embracing the information to improve the quality of care (278).

The qualitative systematic review found that the barriers and facilitators to the use of PROMs in practice emerged within four main themes: practical considerations, attitudes towards the value of the information, methodological considerations and the impact of the feedback on patient care. The review suggested that professional

attitudes to the feedback may be enhanced by engaging with professionals in the planning stage of the intervention, ensuring a high level of transparency around the rationale for the data collection, and also by targeting practical and methodological barriers by using technology to process the information in an efficient manner, standardising data collection processes, providing well-constructed feedback reports and investing in methodological training. Understandably professionals tended to value the information when it proved useful for clinical decision-making. This may be linked to the finding that PROMs were viewed more positively when used as a care management tool for individual patients as a number of studies found that the feedback had the potential to streamline the patient-provider consultation (208-209, 279). Evidence on the perceived value of PROMs as a performance measurement tool was unclear as the two studies that had used PROMs in this manner also provided PROMs feedback at the individual patient level to manage care, and so it was not possible to separate the findings for the different forms of feedback.

Since this review was published, no further publications that meet the review's inclusion criteria have been identified. An additional review was published in 2014 that explored the barriers and facilitators to implementing PROMs in clinical practice in a palliative care setting (280). Although the barriers and facilitators were labelled differently, similar themes emerged to this review namely the themes management/time, education, availability of illness specific instruments, tool specific consideration, motivation/personality/attitudes/beliefs and financing. These are comparable to the practical, attitudinal and methodological themes. However, the financial implications of paying fees for tools were not previously captured. The key facilitators identified were: establishing the role of the coordinator throughout the implementation process, recognising the on-going cognitive and emotional processes

of individuals when implementing change, and providing education to healthcare professionals prior to initiation.

The PROFILE trial

In Chapter 4, the impact of providing surgeons with peer benchmarked PROMs data was examined. Both the quantitative and qualitative reviews highlighted the lack of evidence on the use of PROMs as a performance measurement tool. This is the first randomised controlled trial of PROMs as a performance measurement tool based on the methodology of one of the largest PROMs initiatives— the NHS PROMs Programme. The hypothesis tested was that providing peer benchmarked PROMs data to orthopaedic surgeons would result in better outcomes for patients undergoing hip replacement surgery. The primary outcome for the trial was the post-operative OHS and the secondary outcomes were the HOOS, EQ-5D and the proportion of patients reporting problems after surgery.

This study found no intervention effect, as there was no significant difference between the intervention and control arm. The adjusted effect estimate for the OHS was -0.7 (95% CI -1.9-0.5, P=0.2). Outcomes for patients in both groups improved over the course of the trial, although the differences between pre-feedback and post-feedback outcomes were not statistically significant. Similar findings were observed for the secondary outcome measures.

The improvement in outcomes across the two groups may be attributed to a Hawthorne effect, a contamination effect across surgeons from the intervention to control group, changes in the health system over the course of the study or chance. The lack of a statistically or clinically significant effect could also be explained by the barriers to change identified in the qualitative systematic review such as practical considerations, methodological concerns and attitudes towards the use of PROMs

(142). As the current evidence primarily focuses on the use of PROMs as a patient management tool, there was a need to examine the views and experiences of professionals about the specific use of PROMs as a performance measurement tool. This provided the rationale for the qualitative study.

Qualitative study

In Chapter 5, surgeons' views on and reactions to the peer benchmarked PROMs feedback were explored to examine whether the PROMs information led to changes in provider practice. The findings of this qualitative study provide a unique insight into the use of PROMs as a performance measurement tool and furthermore, add to our understanding of the findings of the PROFILE trial.

Many of the considerations about using PROMs as an individual patient management tool and a performance measurement tool are similar; however this study identified additional and unique challenges to using PROMs as a performance measurement tool. Although there are many common challenges when using PROMs at the individual and aggregated, the use of PROMs as a performance measurement tool also requires adequate expertise and resources to collate the information and perform accurate case-mix adjustment. The four themes identified in the systematic review were also identified in this qualitative study including practical considerations, attitudes towards the value of the data, methodological concerns and the impact of feedback on patient care. A fifth major theme also emerged which captured an issue associated with the surgeon's understanding of the PROMs concept. A similar sub-theme of 'familiarisation', which related to a lack of understanding, was more implicit within the systematic review. The main difference between using PROMs as a patient management tool and a performance measurement tool was in relation to 'methodological considerations'. Surgeons

discussed this at length, expressing a concern about the trustworthiness of data and the threat of using the information as an indicator of their performance.

A typology of surgeons was developed by merging the themes about ‘Attitudes to’ and ‘Impact of’ PROMs information, due to the co-dependency on surgeon’s reactions to the feedback. Three distinct groups of surgeons emerged – Advocates, Converts and Sceptics. Two surgeons were classified as ‘Advocates’ as they expressed a positive attitude to the feedback and stated that the information promoted a self-reflective process on their clinical practice. Four surgeons were classified as ‘Converts’ as they were uncertain about the value of PROMs and this reduced their inclination to use the data. Five surgeons were classified as ‘Sceptics’ as they believed that the PROMs feedback was not clinically useful and claimed that the feedback had no impact on their behaviour.

This study helps us to understand the results of the PROFILE trial. Many surgeons appreciated the feedback, however most considered the information to be insufficient to help identify opportunities for quality improvement. Only two surgeons expressed a positive attitude towards the PROMs feedback and reported that the information promoted them to think about how they deliver care to patients. However, none of the surgeons reported that the information had stimulated an explicit change in their clinical practice. These findings provide possible recommendations for PROMs feedback interventions. Barriers could be reduced by: engaging with professionals from the outset to establish their preferences on the design of the intervention, providing sufficient education and training on the functions of PROMs and interpretation of the information, developing better measures and feedback mechanisms, and building the necessary infrastructure and support to efficiently collect and utilise the PROMs information.

6.2 Implications of findings

6.2.1 Policy and practice

Despite weak evidence of the value of PROMs as a quality improvement tool, the routine collection of PROMs has been implemented in a number of countries including England (3, 72), Australia (105-107), America (103-104), Sweden (104), the Netherlands (108), and interest in the use of PROMs is spreading to other countries (73). Some features of effective policy making include having clearly defined outcomes that the policy aims to achieve, comprehensively reviewing existing evidence and evaluating the impact of the policy (281). The NHS PROMs Programme will be used as an example to explain the implications of our findings on the use of PROMs for policy and practice.

Policy aims and mechanisms of change

The aim of the NHS PROMs Programme was never explicitly stated by the Department of Health (227). The Next Stage Review Interim Report (89), the policy document 'Equity and Excellence: Liberating the NHS' (93) and the consultation document for the Quality and Outcomes Framework (98) specified that the use of PROMs would enhance quality by promoting patient centred care. The Next Stage Review outlined that the collection of PROMs would provide evidence to monitor the effectiveness of care and to link payments to performance (91). The PROMs guidance document outlined that the routine collection of PROMs would promote improvements by benchmarking performance. This information could be used by providers for clinical audit, and by patients, GPs and commissioners to make informed choices (92). Although these references allude to mechanisms by which quality improvements may occur, the objectives of the programme are vague.

Failing to outline clear objectives is indicative of a failure to understand the problem and to develop solutions to combat the problem (281). This process would have led to the development of a model that articulated the causal chain between the policy and its outcomes, and in turn the mechanisms by which quality improvements would be expected to occur. Instead, the Department of Health seems to have taken a simplistic approach to the development of the NHS PROMs Programme—collecting PROMs and publishing the results online, thereby letting the market and behavioural forces linked to Berwick’s model of selection and change pathways take form (100). The problem with such a simplistic approach is that it fails to understand the barriers and facilitators to change.

Example 1- The selection pathway

Selection as a mechanism for improvement does not in itself result in better practice but it can improve outcomes of care by shifting business to the providers with better outcomes (100). For the selection pathway to translate into real quality improvements a number of conditions need to be met: the distribution of performance is relatively stable (i.e. the surgeon’s rank in the distribution is a reliable predictor of his or her future rank and the differences in outcomes are not due to natural variation), patients commissioners and referring clinicians actively seek the information to inform their decisions, and the appropriate market conditions are in place to enable care to shift to the good providers (supply and demand).

Firstly, there is uncertainty about the degree of natural variation in PROMs scores (249). Varagunam *et al.* found that there was no change in the proportion of providers identified as being outliers over the first four years of the NHS PROMs Programme. However, the research did not examine whether the providers identified as being outliers were the same over time (i.e. the proportion could stay the same but

the providers identified as being outliers could differ from one year to the next) (227). Secondly, patients and providers need access to sufficient healthcare information to make rational decisions. However, there is evidence to suggest that patients rarely source publically reported data as: the data does not tell the patient what they really want to know; patients may not be able to correctly interpret results; patients may believe their care is already good; patients think about healthcare at a local level; and patients tend to judge professionals on a more personal level (100, 186, 282). Furthermore, the NHS PROMs Programme has made little attempt to communicate the information appropriately to different audiences. Lastly, in order for the selection pathway to generate better outcomes, there needs to be a large supply of 'good' providers to take the case-load of 'bad' providers and in a situation where the supply of surgeons is finite, there needs to be flexibility within the system to allow high performers to 'scale-up' to meet the increase in demand. Thus, if the selection pathway worked as expected, we should see an improvement in average PROMs scores overtime. A recent publication suggests that this is not happening in practice (227). This example suggests that the NHS PROMs Programme may be necessary but not sufficient for change to occur. In reality, policy makers would either have to make provisions to increase the number of surgeons thereby creating a competitive market, or provide extra capacity for 'good' performers through greater access to resources such as theatres and beds. Successful quality improvement initiatives require a thorough understanding of the policy, the outcomes and the mechanisms by which change can be achieved. This is a problem indicative of the wider quality improvement literature as the evidence on the impact of public reporting is weak (153).

Example 2- The change pathway

The change pathway as a mechanism for improvement aims to stimulate providers to understand the processes of care and to identify what change is required to improve care (100). The level at which feedback is provided may dictate change. Feedback at a trust/hospital level should stimulate top-down change whereas feedback at the surgeon level should stimulate bottom-up change (65). For example, management within a hospital may decide to introduce a pre-assessment clinic to improve the efficacy of care (top-down), alternatively surgeons within a unit may approach management with a proposal to introduce a pre-assessment clinic to reduce risk and improve patient safety (bottom-up). Atul Gawande, a surgeon who is interested in surgical performance, believes that in medicine there is always room for improvement and extremely knowledgeable and skilful doctors need to continuously question what they could do better (186). However, it is well known that change is often resisted in healthcare. Hence, even when changes in surgical care processes have been linked to better outcomes, such is the case with the World Health Organisation surgical checklist (283-284), adherence to such processes remains poor (285). The practical reality is that these processes have proven onerous on providers to implement which explains the slow uptake (274).

At a basic level, failing to identify the aims of a policy leaves providers ambivalent about the purpose of information. Previous research found that providers must understand the function of outcome measurement in order to see the value of the information (286). This message is also echoed in the qualitative systematic review presented in Chapter 3 which highlighted the need to provide transparent objectives for PROMs quality improvement strategies. There was also evidence from the qualitative study presented in this thesis of blurred responsibilities between

clinicians and management. This study identified a possible misconception among participants regarding the objectives of the feedback with some surgeons expressing frustration that it did not provide specific recommendations for improvement. However, from the policy maker's point of view, the feedback is supposed to stimulate the professionals to identify areas for improvement. This finding highlights a gap in the perceived objectives of the feedback between clinicians and policy makers, and therefore responsibilities should be explicitly outlined.

This section highlights why it is important to outline the aims and objective of the policy, and the mechanisms by which improvements should occur.

Evidence and implementation strategy

The NHS PROMs Programme was not based on strong evidence of effectiveness. The quantitative systematic review in Chapter 2 found that there is little or no evidence to support this approach. Also, the initiative was not based on a well formulated implementation strategy as it did not take into consideration the barriers and facilitators to change. Although the NHS PROMs Programme ran a pilot project, this concentrated on the feasibility of data collection rather than the effectiveness of the information (1). One could argue that the programme was introduced by the Department of Health on a 'trying it out' basis for the four elective procedures enabling the effectiveness to be tested in advance of wider implementation to other procedures or conditions (287). However, implementing a policy without taking into consideration the barriers to change could in fact damage the reputation of PROMs by increasing the bureaucratic burden with little positive impact on care (225) and furthermore by attributing the failure to the measurement model rather than the implementation strategy itself (249).

Varagunam *et. al* suggested that the lack of impact of the NHS PROMs Programme on patient outcomes may be explained by the implementation strategy, in particular the feedback strategy and the lack of support available to advise providers on what action should be taken to improve care (227). These claims tie in with the findings of the qualitative review in Chapter 3 and the results of interviews with surgeons in Chapter 5. The review identified that professionals want more sophisticated feedback clearly depicting what constitutes a clinically important change (141). Currently, the NHS PROMs Programme information is presented in unwieldy Excel spread sheets on the Information Centre's web site. To the untrained eye, this information is extremely hard to find, to navigate and to interpret. Therefore, the value of such a potentially worthwhile initiative may be lost due to the lack of a relatively small investment in communication and consultation (269). In addition, the qualitative interviews with surgeons identified that the PROMs data alone were insufficient to provide explanations of poor performance, and surgeons did not have the training or support to examine the reasons for differences in outcomes so the feedback had little relevance in practice. PROMs data may be more effectively used if they are appropriately fed back to providers and accompanied by a level of support. It is important to engage with providers to identify a common goal for measurement, to educate and train providers to enable them to use the information, and to assist professionals when undertaking further audit and research activities.

6.2.2 Research

As the use of PROMs as a quality improvement tool is a relatively novel approach, it is a fertile ground for research. Many interesting research questions have emerged from this thesis. These have been categorised into five sections: how PROMs may work (as a tool for quality improvement), what to measure, is the data of good

quality, what are the most appropriate methods for analysing and interpreting the data, and how to get the most out of PROMs? Future research on PROMs will help advance the science of outcome measurement in general (288).

How PROMs may work?

One clear message emerging from this research is that we do not fully understand the mechanisms by which PROMs may lead to change. This is an important area for future research as it is necessary to gain an understanding of how this information may lead to changes in practice (49, 51, 82, 147). Greenhalgh *et al.* are currently undertaking a realist synthesis which builds on the findings of our systematic review on the use of PROMs given the different functions of the data. This reviews aims to understand by what means and in what circumstances PROMs feedback leads to intended service improvements (289). It will advance our current understanding by evaluating the evidence in light of a comprehensive set of theories. Logic models will be used to build different ideas and assumptions about how PROMs feedback is supposed to impact on practice (289).

This process may enhance thinking on the particular set of circumstances required to promote change. It is important to advance the implementation of PROMs quality improvement interventions, promoting researchers and policy makers to use more simple logic by outlining clear objectives, examining theories of change, detailing causal pathways as well as predicting expected mechanisms and barriers to change (290). Future research should build on our work by testing hypotheses which have been informed by theory, particularly in areas which are poorly understood in the wider quality improvement literature such as the use of audit and feedback when linked to benchmarking, public release of information and pay for performance (30, 41, 282, 291-292).

What measure to use?

The second area for further investigation highlighted by this research is the on-going need to advance and refine measurement. There are over one thousand PROMs available so it is not always clear which is the most appropriate measure to use, particularly given that PROMs can be used for a number of different purposes (293). At this point, it is important to discuss what constitutes a good measure.

Our lives are full of instruments that help us to quantify and understand elements around us such as clocks, weighing scales, rulers, and thermometers. In order for these instruments to be useful, they need to be accurate. Accurate measurement requires that the instrument is reliable by giving the same reading if something is measured twice, e.g. similar readings should be displayed if one stands on a weighing scale twice. Good measurement also requires that the instrument is valid by measuring what one intends to measure, e.g. a car that measures kilometres when one wants to measure miles. Lastly, a good measure should be responsive to detect a meaningful quantity of change, e.g. a measuring tape that displays centimetres when one wants to measure millimetres. These attributes also apply to measurement tools used to quantify social and psychological variables like pain and function.

Psychometrics is the study of methods for measuring social and psychological variables. The reliability, validity and responsiveness are often referred to as measurement properties (294). There are many different terms used to define measurement properties. Therefore, the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) initiative was developed with the aim of improving the selection of PROMs by clarifying definitions and meanings of properties through a consensus process. The final taxonomy included three domains for consideration: reliability (internal consistency, reliability and measurement

error), validity (content validity, construct validity and criterion validity) and responsiveness. Interpretability was also identified as an important element to consider, although this was not categorised as a measurement property (294). It is important to be aware of the purpose for which the measure was developed and psychometrically tested, and to acknowledge the intended use of the measure to establish whether the measurement properties still hold (293). The Oxford Hip Score (OHS), for example, was developed to assess the clinical effectiveness of hip replacement surgery for groups of patients (295). In this instance, the responsiveness of the measure to detect change over time would have been of primary importance. In the context of performance measurement, it is important that the tool can discriminate between providers (293, 295).

This presents the concern that different psychometric issues are raised depending on the aim of the data collection. Psychometric theory argues that PROMs are validated for a particular purpose (293). Many of the commonly used PROMs were developed to compare groups of patients. The OHS has been psychometrically tested to detect change in pain and function before and after a hip replacement operation (231, 296). However, this measure has subsequently been used for different purposes (2-3). The NHS has used the OHS as a tool to detect differences in quality between healthcare providers, without formally testing this function (3, 228). The National Quality Forum (NQF) emphasise that a psychometrically sound PROM does not directly translate into a good measure of performance (228).

Firstly, the magnitude of change in the OHS between pre- and post-surgery is completely different to the magnitude of discrepancies in scores across providers. One can expect an average change score of 20 points in the OHS between pre- and post-surgery whereas the PROFILE trial found that the difference in change scores

between the extreme performers was at most 7 points. Therefore, we cannot assume without the appropriate testing that the OHS is responsive to detect differences in quality between providers. This scenario is conceptually similar to the analogy of measuring centimetres and millimetres. Secondly, there is the additional concern when using PROMs as measures of performance in the ability of the tool to capture the full range of measurement (228). Many measures, including the OHS, have floor or ceiling effects (297). A floor effect is conceptually similar to a ruler which is missing the first 5 centimetres and a ceiling effect is conceptually similar to a ruler which can only measure up to 20 centimetres. This poses a problem when in fact one intends to measure a construct which ranges beyond these values. In essence, when a measurement tool is subject to these floor and ceiling effects, we are not capturing the full range of the concept being measured. The OHS is subject to ceiling effects and the significance of this in the context of performance measurement should not be underestimated as it may lead to the inappropriate labelling of a provider's performance (228). These measurement considerations have different implications for policy as using inadequate tools to drive decisions such as pay for performance will lead to an inappropriate allocation of resources across providers (3).

The argument that different psychometric issues are raised by the multiple uses of PROMs can be further emphasised when using a measurement tool, which is psychometrically tested at the group level, to assess individual patient scores. Once again, developing policies based on individual level measurement with inappropriate tools may have very serious consequences (148). To illustrate, envision the impact of using an inappropriate measure to help prioritise patients for eligibility for particular healthcare interventions. This may result in eligible candidates being refused treatment, such as hip replacement surgery, on the basis of a poor measurement tool.

In this context, developing and testing tools through modern psychometric methods such as Rasch analysis would offer great benefits by enabling the generation of individual standard errors to assess the accuracy of measurement at this level (148, 298).

The COSMIN group also developed a checklist to evaluate measures (294). The checklist includes a section on the statistical methods used to develop the measure. This methods used is an important consideration when choosing PROMs and for the future development of PROMs. Traditional psychometric methods are underpinned by a theory called Classical Test Theory (CTT). This evaluates measures in terms of the psychometric properties previously mentioned. CTT has been widely used in outcome measurement and many of the popularly used measures have been developed through this method (299). CTT focuses on test level information which looks at the sum of responses to items. It uses ordinal scales and assumes item equivalence meaning that each item contributes equally to the final score irrespective of how well they correlate with the final score (298, 300). Modern psychometric properties are underpinned by a theory called Latent Trait Theory (LTT) which primarily refers to two methods called Item Response Theory (IRT) and Rasch analysis. These methods can offer additional benefits to the development of measures as they have the potential to improve the accuracy of measurement. LTT focuses on item-level information meaning that they build on the relationship between a person's answer on an item (e.g. climbing stairs or walking a block) and the score of the concept being measured (e.g. physical functioning) (300). IRT prioritise the data (finding a model that best explains the data) and Rasch prioritises the model (if data does not fit, it seeks to understand why). Most fundamentally, these approaches develop interval scales which are more accurate and enable the

determination of item fit and difficulty. The concepts of item fit and difficulty offer benefits to scale development. These determine the items which capture a particular range of difficulty and can reduce the number of items on a scale according to how they fit within the range. The use of LTT offers enormous potential for the development of PROMs. Most importantly, more precise measurement should translate into a reduction in the burden of data collection in practice (223, 270, 298).

In summary, as many PROMs are validated for a particular purpose caution should be applied when planning to use the tools for multiple purposes. This is because many tools have been developed through CTT methods which have a number of limitations including: the data generated are ordinal, scores for persons and samples are scale-dependent, scale properties are sample-dependent and the data are only suitable for group studies, and not individual patient measurement (148). Modern psychometric methods can offer a solution to the development of more accurate PROMs that are more suitable if the measure is going to be used for multiple purposes. Firstly, they have the ability to construct interval level scaling as opposed to ordinal level. Therefore, units on the scale are standardised which makes the interpretation of change scores more meaningful. Secondly, they enable the generation of individual level standard errors, so the measures can be analysed at the individual patient level as well at the group level. This ensures that PROMs can be reliably used at different levels and for different purposes. Lastly, scales are developed by understanding the relationship between the construct being measured and the items in the scale. Therefore, they provide item estimates that are free from a sample distribution and person estimates that are free from a scale distribution. This means that a subset of items from the scales can be used, which are comparable to scores derived from a different set of items. This reduces respondent burden and

improves accuracy by ensuring items are targeted to the sample being measured (148, 223, 298).

Data quality and integrity

The third area for future research is the evaluation of the quality of data. This is a pertinent issue when using PROMs as performance measurement tools. High quality data support the provision of effective decision making (301). To build professionals' confidence in the information, it is important to demonstrate data quality. Incorrect or incomplete data are a major concern for professionals (302). The qualitative research in Chapter 5 identified that surgeons were particularly sceptical about the quality of patient-reported data. Surgeons were concerned about the risk of patients completing the questionnaire incorrectly, especially those with co-morbidities and poor literacy. They were also worried about the potential for providers to manipulate scores by failing to recruit patients who may be more likely to have a poor outcome. The importance of data quality cannot be stressed enough. The National Surgical Quality Improvement programme attributes part of its success to the efforts dedicated to ensuring data integrity (274).

There are two main concerns in relation to data quality- the timeliness and the completeness of the data. The timeliness of PROMs can be a problem as there is often a lag between the event, the outcome and the feedback. This is a contentious issue which could be improved by the effective use of technology enabling more efficient data collection and instant feedback. However, validation of electronic systems to collect PROMs data and the implications of using different modes of data collection are important considerations for future research (303-304). The completeness of the data is the most serious threat to the value of this information. The completeness of data is determined by recruitment rates and response rates.

Developing data collection protocols, training data collectors and standardising data collection processes across different sites can improve recruitment. However, continuous investigation into recruitment rates across providers is necessary. Similar to the response rates in the PROFILE trial, the North East Quality Observatory System in England found that the NHS PROMs data represented less than 50% of actual activity (241). The problem is that it is not possible to determine if the recruitment rates are linked to a selection bias as some data collectors may inadvertently cherry pick patients by encouraging healthier patients to complete the questionnaire. This makes it difficult to report with confidence that variations in outcomes across providers are accurate. The NHS PROMs Programme are currently attempting to incentivise better recruitment by linking provider's payments to their recruitment rates (98). Another potential source of bias in patient recruitment is the exclusion of those that cannot self-complete a questionnaire because of literacy (242, 305) and language comprehension issues (306). Translating a measure into multiple languages is problematic as it can be difficult to ensure the correct translation and cultural adaptation of measures, and requires revalidation of the tool. However, if the instruments are not available in multiple languages and are not user friendly, minority groups and patients with poor literacy are excluded which also may introduce biased estimates across providers (220, 307). The second issue to consider are patient response rates. Response rates for the PROFILE trial were high. However, Hutchings *et al.* found that response rates in England differed between 30-100% across healthcare providers. The evidence suggests that non-responders tend to have a poorer pre-operative quality of life, indicating that rates of non-response need to be considered when comparing the performance of providers (308-309).

Continued efforts to monitor and incentivise better data quality will inevitably generate greater confidence in the value of the PROMs data.

Analysis and interpretation

The fourth area for future research involves exploring analysis and interpretation methods. Ensuring accurate case-mix adjustment is vital when using PROMs to compare providers. Patient profiles vary across providers due to differences in populations surrounding clinics or hospitals. Some providers treat a riskier case-mix of patients so a significant amount of the variation between providers can be explained by patient characteristics. Therefore, to accurately compare professionals the analysis has to adjust for these differences (59). Professionals often do not trust the accuracy of case-mix adjustment methods (150) and in fact, they have some justification for such scepticism as case-mix adjustment is far from a perfect science. To adjust for confounding variables, firstly it is necessary to understand which variables predict the outcome outside the control of the provider, and secondly it is necessary to have access to these data to enable adjustment. Other important factors that may influence performance beyond patient characteristics include the institutional structure (size, equipment, staffing levels, teaching status) and intervening variables (culture, stress, availability of staff) (3, 60, 310). However, significant questions remain regarding which variables should be included in the model and which analytical technique is the most appropriate to use (288). This is an area which is subject to on-going research. The NHS PROMs Programme published guidelines in 2012 on suggested case-mix adjustment models (232), and updated methods in 2013 and 2014 in light of feedback from clinicians and other stakeholders (250).

PROMs produce data that are inherently hard to interpret as instruments differ in respect to items, response options and approaches to aggregation (190, 311). Interpretation translates data into familiar recognisable terms (312) and is defined as ‘the degree to which one can assign qualitative meaning— that is, clinical or commonly understood connotations— to quantitative scores’ (313). PROMs are based on constructs that cannot be objectively measured so meaning must be gained through indirect measurement. For example, functional status can be measured through one’s ability to perform daily tasks such as climbing a stairs and doing housework. Interpreting the significance of a change in score can be difficult to comprehend as many commonly used measures have ordinal scales which do not have precise units of measurement (223, 311). This creates a challenge when trying to establish the clinical meaning of results (220). In order to translate the data into meaningful terms, intuitive benchmarks are required to interpret the data (190). These methods tend to focus on establishing minimally and clinically important differences (190, 314). However, the use of LTT also can facilitate the transformation of PROMs into interval scales, as well as enabling items to be linked to specific scores, offering benefits for interpretation (190, 223). Continued research is necessary to help make sense of PROMs information and to identify how best to present this information to different stakeholders (268-269).

‘Actionability’ of the data

The fifth area for future research involves understanding the causes of variation in outcomes. PROMs feedback provides evidence on differences in outcomes between providers but it does not offer knowledge of the underlying reasons for these variations. The correlation between process and outcome measures remains poorly understood so a major frustration for professionals with the use of PROMs as

performance measurement tools is the ‘actionability’ of the data (315). The ‘actionability’ refers to the extent to which the information identifies solutions for improvement. The causal pathway between processes of care and the outcomes of care is complicated making it difficult to establish why the variation actually exists (60).

Future research should perform case studies of top performers to examine what they are doing differently compared to the poor performers. Linking outcomes data to administrative databases may enable more sophisticated analysis to uncover relationships between clinical parameters, such as surgical techniques and approaches, and outcomes. Another interesting area for exploration is the relationship between patient satisfaction, expectations and outcomes. A recent review suggests that there is a positive correlation between satisfaction and PROMs, but the causality of this relationship is unknown. In essence, we do not know if a better healthcare experience leads to a better perception of one’s outcome or if a better healthcare experience provides patients with the ability to manage their healthcare issue better, and hence promotes to better outcomes (271).

Getting the most out of PROMs

In order to determine the benefit of using PROMs as a quality improvement tool, further quantitative and qualitative research is required to determine the impact of using PROMs on patient care.

One could argue that sufficient time has not elapsed since the NHS PROMs Programme commenced for structural and process changes to filter through to patient outcomes, so on-going time-series analysis should be undertaken (58). The Department of Health is currently running pilot projects to extend the PROMs Programme to a wider range of conditions in the NHS including: mental health,

cancer care and long-term conditions (asthma, COPD, diabetes, epilepsy, heart failure and stroke). The extension of the Programme brings additional methodological challenges as these conditions do not have specific intervention points to carry out the ‘before and after’ type approach currently employed by the NHS PROMs Programme (3). If continuing to use PROMs as a performance measurement tool, one important aspect to consider is the variability between providers. Lyratzopoulos suggests that investigating provider level heterogeneity could help prioritise healthcare improvement efforts by identifying the conditions associated with greater potential for quality improvement. Performance measurement may be best targeted to areas where there is high variability ($ICC > 10$) between providers (316). Low variation and high baseline performance signals that quality of care is of a high standard so the scope for improvement may be too small to justify performance monitoring. However, this should be interpreted with caution as low baseline performance and low variation between providers may indicate that outcomes are poor across all providers and improvement may be possible across the board. A potentially useful scoping exercise would be to explore variation between providers across different conditions or procedures. It is important to acknowledge that the variation becomes more complex when focusing on conditions that have an unpredictable or a flaring nature so once-off measurement may not be an accurate reflection of performance (317). Consequently, further effort is required to establish the optimal time point to assess patient-reported outcomes for different healthcare conditions (318).

The most concerning issue when employing performance monitoring is the potential for unintended consequences for providers and patients. There is an argument that performance monitoring can slow change as providers may focus efforts on

improving the indicators under observation at the expense of other aspects of care (302). Doran *et. al* found that improvements occurred across indicators within the UK's Quality and Outcomes Framework, however these were achieved at the expense of aspects of care that were not incentivised (319). Performance monitoring may also affect access and equity, promoting professionals to treat the least risky patients to improve their ranking (150, 320). Examining the case-mix of professionals before and after the introduction of these quality improvement programmes may help to establish the extent to which this occurs in practice (227). Little research has been dedicated to examining unintended consequences of the NHS PROMs Programme. The potential for 'gaming' may become more pronounced if the decision to link performance to payment is implemented. Furthermore, it has been recommended that performance indicators in Quality and Outcomes Framework should be replaced if they are not proving to be effective (321). The outstanding question is whether the NHS PROMs Programme should adopt a similar approach? However, there is a danger that decisions will be made without addressing some of the underlying practical and methodological issues identified in this thesis, sending out a signal that using PROMs is not effective rather than identifying problems with the implementation of the strategy.

The extent to which professionals are willing and able to implement change needs further exploration. The qualitative study in Chapter 5 found that participants varied in their understanding of the concept of PROMs. This finding may be more applicable in Ireland where the collection of performance data is not a common feature of the system. Exploring providers understanding of what PROMs are would be an interesting research study in the NHS where PROMs are now a strong feature of care (322). In addition, further qualitative research could help to unpick the causes

of resistance to change and the level to which there is a conflict in objectives among different stakeholders.

The use of PROMs for quality improvement purposes is only one facet to the purpose of measuring patient-reported outcomes. The use of PROMs in clinical, economic, health services research and general population health assessment is likely to grow. One exciting development in this field is the creation of item banks and computerized adaptive testing which can reduce the burden of data collection by targeting the appropriate questions to specific patient groups (323). Furthermore, advances in healthcare information systems will enable the integration of PROMs data into electronic health records, expanding opportunities for using the data for multiple purposes (324).

6.2.3 Implications of the findings in Ireland

The findings from this research suggest that the evidence base is currently too weak to recommend the routine collection of PROMs to monitor the performance of providers in Ireland. In a country where outcome measurement is not commonly applied, knowledge and familiarity of PROMs is low, and the IT infrastructure within the system is limited; the likelihood of resistance towards a PROMs policy would be high. It is possible that the level of resistance would be further accentuated if the information was publically reported as is the practice in the NHS. Even if the documented conceptual, methodological and practical challenges identified in this thesis were addressed, the current structure and funding of the health system in Ireland would not facilitate an incentive to improve. For example, applying the hypothetical mechanisms of change as explained earlier in this section through Berwick's model of 'selection' requires specific conditions, which are not currently present in Ireland (100). The demand for care in Ireland is far greater than the supply

of care. Patients, particularly in the public system, are often forced to wait for years to get access to a specialist consultant and to the appropriate healthcare intervention. At this point, patients are so anxious to get care that the selection of a particular provider is not a priority (125). More importantly, for 'selection' to work as a mechanism of change, the poorly performing providers need to be pushed out of the market by offering patients with a choice of higher performing providers. Again, this is unlikely to occur in Ireland as the current system is currently burdened by a recruitment crisis (125). However, the 'change' pathway may offer more favourable opportunities for improvement (100). The qualitative evidence identified a desire by professionals to continuously develop their competencies. The uptake of PROMs in practice may be improved by building the infrastructure to effectively collect and disseminate the information, as well as developing knowledge through appropriate education and training programmes. By providing appropriate systems and supports, a bottom-up appreciation of the information derived from PROMs may be ensued. Therefore, building these capabilities within our system should naturally drive a desire for the use of PROMs in practice. This in turn may promote the research agenda for PROMs, helping to develop the measures methodologically.

6.3 Limitations of the research

Every attempt was made to produce high quality research for this thesis. However there are a number of limitations. First, one reviewer performed the initial screening and study selection for the systematic reviews, and although reference searching was undertaken to reduce the likelihood of missing appropriate studies, there is a chance that relevant literature was not identified. Second, cluster randomised controlled trials can have limitations compared to individually randomised controlled trials

(325-326). Imbalance is more likely to occur as there are relatively few clusters to randomise compared to the number of individuals in the trial. We employed stratified randomisation to increase the likelihood that arms would be balanced with respect to important predictor variables and this resulted in similar surgeon and patient characteristics across groups. Cluster randomised trials are also less powerful than individually randomised trials with the same number of participants because participants within clusters are not independent. We inflated the sample size accordingly to achieve sufficient power. However, this impacts on the feasibility of running a trial in practice. We had difficulty recruiting patients across a number of sites, which was primarily explained by the rotation of staff recruiting patients and a reluctance to take on the additional workload. We used techniques to improve recruitment such as sending monthly newsletters to data collectors and surgeons comparing recruitment rates across sites (Appendix 15), contacting data collectors each week to get an update on recruitment and introducing monthly prizes to incentivise better recruitment. Another limitation of our study design was that surgeons within eight hospitals were randomised to control and intervention arms thus creating an opportunity for a contamination effect of the intervention. We considered alternative designs to prevent this effect. Randomising by hospital was not an option as a number of surgeons work across multiple sites, particularly in Dublin, and the number of clusters would have been too small to randomise at a regional level. Third, while this research focused on the impact of PROMs feedback on improving patient outcomes, other indicators of quality such as communication and patient satisfaction may be appropriate targets for improvement. Fourth, the PROMs used were based on the tools used by the NHS PROMs Programme, so the impact of the information may be a reflection of the rigour of the current measures

available. Finally, the qualitative study only explored surgeon's experiences and did not elicit the views of patients, other healthcare professionals or healthcare managers about the value of PROMs. Therefore, the findings will only represent part of the overall picture in terms of attitudes to the value and challenges of PROMs in practice.

6.4 Conclusion

Although the use of PROMs as a quality improvement tool is gaining interest internationally, little effort has been made to understand the mechanisms by which this information may lead to improvements in the quality of care. This research demonstrated that peer benchmarked PROMs feedback had minimal impact on the behaviour of surgeons. The qualitative study identified the reasons for the observed reluctance of providers to embrace PROMs as conceptual, methodological and practical factors. Methods to address potential barriers to change include consulting with professionals at the developmental stage of a feedback initiative, communicating with professionals about the objectives of the data collection, educating professionals on the properties and interpretation of the data, and supporting professionals in using the information to identify areas for improvement.

References

1. Browne J, Jamieson L, Lewsey J, van der Meulen J. 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.
2. 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.
3. Devlin N, Appleby J. Getting the most out of PROMs: putting health outcomes at the heart of NHS decision-making. London: King'd Fund 2010.
4. Institute of Medicine, editor. Crossing the Quality Chasm: A New Health System for the 21st Century Washington, D.C: National Academy Press; 2001.
5. Eagle CJ, Davies JM. Current models of "quality"--an introduction for anaesthetists. Canadian journal of anaesthesia = Journal canadien d'anesthesie. 1993 Sep;40(9):851-62.
6. Marjoua Y, Bozic KJ. Brief history of quality movement in US healthcare. Current reviews in musculoskeletal medicine. 2012 Dec;5(4):265-73.
7. Institute of Medicine. To Err Is Human: Building a Safer Health System. Washington, D.C.: National Academy Press; 2000.
8. Sturmburg JP, Martin CM. Complexity and health--yesterday's traditions, tomorrow's future. J Eval Clin Pract. 2009 Jun;15(3):543-8.
9. Bilawka E, Craig BJ. Quality assurance in health care: past, present and future. Int J Dent Hyg. 2003 Aug;1(3):159-68.
10. Guth KA, Kleiner B. Quality assurance in the health care industry. Journal of health care finance. 2005 Spring;31(3):33-40.
11. Spiegelhalter DJ. Surgical Audit: Statistical Lessons from Nightingale and Codman. 1999.
12. Chassin MR, Loeb JM. The ongoing quality improvement journey: next stop, high reliability. Health Aff (Millwood). 2011 Apr;30(4):559-68.
13. Institute of Medicine. Institute of Medicine. America 2014; Available from: <http://www.iom.edu/>.
14. The Agency for Healthcare Research and Quality. The Agency for Healthcare Research and Quality America 2014; Available from: <http://www.ahrq.gov/>.
15. The National Committee for Quality Assurance. The National Committee for Quality Assurance America 2014; Available from: <http://www.ncqa.org/>.
16. Institute for healthcare improvement. Institute for healthcare improvement. 2014; Available from: <http://www.ihl.org/Pages/default.aspx>.
17. ISQua. The International Society for Quality in Health Care. 2014; Available from: <http://www.isqua.org/home>.
18. Care Quality Commission. Care Quality Commission. England 2014; Available from: <http://www.cqc.org.uk/>.
19. Canadian Foundation for Healthcare Improvement. Canadian Foundation for Healthcare Improvement. Canada 2014; Available from: <http://www.cfhi-fcass.ca/Home.aspx>.
20. Australian Commission on Safety and Quality in Health Care in Australia. Australian Commission on Safety and Quality in Health Care in Australia. Australia 2014; Available from: <http://www.safetyandquality.gov.au/>.

21. HIQA. Health Information and Quality Authority. Ireland 2014; Available from: <http://www.hiqa.ie/>.
22. Chow A, Mayer EK, Darzi AW, Athanasiou T. Patient-reported outcome measures: the importance of patient satisfaction in surgery. *Surgery*. 2009 Sep;146(3):435-43.
23. Weiss KB, Wagner R. Performance measurement through audit, feedback, and profiling as tools for improving clinical care. *Chest*. 2000 Aug;118(2 Suppl):53S-8S.
24. Kunkel S, Rosenqvist U, Westerling R. The structure of quality systems is important to the process and outcome, an empirical study of 386 hospital departments in Sweden. *BMC Health Serv Res*. 2007;7:104.
25. The Health Foundation. Quality improvement made simple- What every board should know about healthcare quality improvement. 2013.
26. Shojania KG, McDonald KM, Wachter RM, Owens DK. Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies. In: Markowitz AJ, editor. U.S.: Agency for Healthcare Research and Quality; 2004.
27. Cheung A, Weir M, Mayhew A, Kozloff N, Brown K, Grimshaw J. Overview of systematic reviews of the effectiveness of reminders in improving healthcare professional behavior. *Systematic reviews*. 2012;1:36.
28. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. *Int J Med Inform*. 2010 Nov;79(11):736-71.
29. Cappuccio FP, Kerry SM, Forbes L, Donald A. Blood pressure control by home monitoring: meta-analysis of randomised trials. *BMJ*. 2004 Jul 17;329(7458):145.
30. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev*. 2012;6:CD000259.
31. Forsetlund L, Bjorndal A, Rashidian A, Jamtvedt G, O'Brien MA, Wolf F, et al. Continuing education meetings and workshops: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*. 2009(2):CD003030.
32. Boyde M, Turner C, Thompson DR, Stewart S. Educational interventions for patients with heart failure: a systematic review of randomized controlled trials. *J Cardiovasc Nurs*. 2011 Jul-Aug;26(4):E27-35.
33. Mason J, Khunti K, Stone M, Farooqi A, Carr S. Educational interventions in kidney disease care: a systematic review of randomized trials. *American journal of kidney diseases : the official journal of the National Kidney Foundation*. 2008 Jun;51(6):933-51.
34. Monninkhof E, van der Valk P, van der Palen J, van Herwaarden C, Partridge MR, Zielhuis G. Self-management education for patients with chronic obstructive pulmonary disease: a systematic review. *Thorax*. 2003 May;58(5):394-8.
35. Foster G, Taylor SJ, Eldridge SE, Ramsay J, Griffiths CJ. Self-management education programmes by lay leaders for people with chronic conditions. *Cochrane Database Syst Rev*. 2007(4):CD005108.
36. Vervloet M, Linn AJ, van Weert JC, de Bakker DH, Bouvy ML, van Dijk L. The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: a systematic review of the literature. *J Am Med Assoc*. 2012 Sep-Oct;19(5):696-704.
37. Bamberger SG, Vinding AL, Larsen A, Nielsen P, Fonager K, Nielsen RN, et al. Impact of organisational change on mental health: a systematic review. *Occupational and environmental medicine*. 2012 Aug;69(8):592-8.

38. Andersen H, Rovik KA, Ingebrigtsen T. Lean thinking in hospitals: is there a cure for the absence of evidence? A systematic review of reviews. *BMJ Open*. 2014;4(1):e003873.
39. Elkhuizen SG, Limburg M, Bakker PJ, Klazinga NS. Evidence-based re-engineering: re-engineering the evidence--a systematic review of the literature on business process redesign (BPR) in hospital care. *Int J Health Care Qual Assur Inc Leadersh Health Serv*. 2006;19(6-7):477-99.
40. Parmelli E, Flodgren G, Schaafsma ME, Baillie N, Beyer FR, Eccles MP. The effectiveness of strategies to change organisational culture to improve healthcare performance. *Cochrane Database Syst Rev*. 2011(1):CD008315.
41. Eijkenaar F, Emmert M, Scheppach M, Schoffski O. Effects of pay for performance in health care: a systematic review of systematic reviews. *Health Policy*. 2013 May;110(2-3):115-30.
42. Sutherland K, Leatherman S. Regulation and quality improvement- A review of the evidence: The Health Foundation 2009.
43. Bero LA, Grilli R, Grimshaw JM, Harvey E, Oxman AD, Thomson MA. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. The Cochrane Effective Practice and Organization of Care Review Group. *BMJ*. 1998 Aug 15;317(7156):465-8.
44. Davis DA, Thomson MA, Oxman AD, Haynes RB. Changing physician performance. A systematic review of the effect of continuing medical education strategies. *JAMA*. 1995 Sep 6;274(9):700-5.
45. Grimshaw JM, Shirran L, Thomas R, Mowatt G, Fraser C, Bero L, et al. Changing provider behavior: an overview of systematic reviews of interventions. *Med Care*. 2001 Aug;39(8 Suppl 2):II2-45.
46. O'Brien MA, Rogers S, Jamtvedt G, Oxman AD, Odgaard-Jensen J, Kristoffersen DT, et al. Educational outreach visits: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*. 2007(4):CD000409.
47. Oxman AD, Thomson MA, Davis DA, Haynes RB. No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. *CMAJ*. 1995 Nov 15;153(10):1423-31.
48. Jaspers MW, Smeulders M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *J Am Med Inform Assoc*. 2011 May 1;18(3):327-34.
49. Colquhoun HL, Brehaut JC, Sales A, Ivers N, Grimshaw J, Michie S, et al. A systematic review of the use of theory in randomized controlled trials of audit and feedback. *Implement Sci*. 2013;8:66.
50. Eccles M, Grimshaw J, Walker A, Johnston M, Pitts N. Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. *J Clin Epidemiol*. 2005 Feb;58(2):107-12.
51. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement Sci*. 2011;6:42.
52. Cochrane Effective Practice and Organisation of Care Review Group. EPOC-Data Collection Checklist. The Cochrane Collaboration; 2014; Available from: <http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/datacollectionchecklist.pdf>.

53. Woolf S, Schunemann HJ, Eccles MP, Grimshaw JM, Shekelle P. Developing clinical practice guidelines: types of evidence and outcomes; values and economics, synthesis, grading, and presentation and deriving recommendations. *Implement Sci.* 2012;7:61.
54. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA.* 2006 Sep 6;296(9):1094-102.
55. Smith WR. Evidence for the effectiveness of techniques To change physician behavior. *Chest.* 2000 Aug;118(2 Suppl):8S-17S.
56. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Does telling people what they have been doing change what they do? A systematic review of the effects of audit and feedback. *Qual Saf Health Care.* 2006 Dec;15(6):433-6.
57. Hysong SJ. Meta-analysis: audit and feedback features impact effectiveness on care quality. *Med Care.* 2009 Mar;47(3):356-63.
58. Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Med Teach.* 2006 Mar;28(2):117-28.
59. Mant J. Process versus outcome indicators in the assessment of quality of health care. *Int J Qual Health Care.* 2001 Dec;13(6):475-80.
60. Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet.* 2004 Apr 3;363(9415):1147-54.
61. McCartney P, Macdowall W, Thorogood M. A randomised controlled trial of feedback to general practitioners of their prophylactic aspirin prescribing. *BMJ.* 1997 Jul 5;315(7099):35-6.
62. Epstein AM. The outcomes movement--will it get us where we want to go? *N Engl J Med.* 1990 Jul 26;323(4):266-70.
63. Bauchner H, Simpson L, Chessare J. Changing physician behaviour. *Arch Dis Child.* 2001 Jun;84(6):459-62.
64. O'Connor GT, Plume SK, Olmstead EM, Morton JR, Maloney CT, Nugent WC, et al. A regional intervention to improve the hospital mortality associated with coronary artery bypass graft surgery. The Northern New England Cardiovascular Disease Study Group. *JAMA.* 1996 Mar 20;275(11):841-6.
65. Bridgewater B, Keogh B. Surgical "league tables": ischaemic heart disease. *Heart.* 2008 Jul;94(7):936-42.
66. Bridgewater B, Grayson AD, Brooks N, Grotte G, Fabri BM, Au J, et al. Has the publication of cardiac surgery outcome data been associated with changes in practice in northwest England: an analysis of 25,730 patients undergoing CABG surgery under 30 surgeons over eight years. *Heart.* 2007 Jun;93(6):744-8.
67. Epstein AJ. Do cardiac surgery report cards reduce mortality? Assessing the evidence. *Med Care Res Rev.* 2006 Aug;63(4):403-26.
68. Hannan EL, Cozzens K, King SB, 3rd, Walford G, Shah NR. The New York State cardiac registries: history, contributions, limitations, and lessons for future efforts to assess and publicly report healthcare outcomes. *Journal of the American College of Cardiology.* 2012 Jun 19;59(25):2309-16.
69. White KL. Improved medical care statistics and the health services system. *Public Health Rep.* 1967 Oct;82(10):847-54.
70. Health Services Research Group. Outcomes and the management of health care. *CMAJ.* 1992 Dec 15;147(12):1775-80.

71. McGrail K, Bryan S, Davis J. Let's all go to the PROM: the case for routine patient-reported outcome measurement in Canadian healthcare. *Healthcare Papers*. 2011;11(4):8-18; discussion 55-8.
72. Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 2013;346:f167.
73. Canadian Institute of Public Health. *Health Outcomes of Care: An Idea Whose Time Has Come*. Ottawa, Ontario 2012.
74. Balanda KP, Barron S, Fahy L, McLaughlin A. *Making Chronic Conditions Count: Hypertension, Stroke, Coronary Heart Disease, Diabetes. A systematic approach to estimating and forecasting population prevalence on the island of Ireland*. Dublin: Institute of Public Health in Ireland 2010.
75. Fayers PM, Machin D. *Quality of Life- The assessment, analysis and interpretation of patient-reported outcomes*. West Sussex, England: Wiley; 2007.
76. 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-iv, 1-74.
77. Lipscomb J, Gotay CC, Snyder CF. Patient-reported outcomes in cancer: a review of recent research and policy initiatives. *CA Cancer J Clin*. 2007 Sep-Oct;57(5):278-300.
78. Bendtsen P, Leijon M, Sofie Sommer A, Kristenson M. Measuring health-related quality of life in patients with chronic obstructive pulmonary disease in a routine hospital setting: feasibility and perceived value. *Health Qual Life Outcomes*. 2003;1:5.
79. Dinan MA, Compton KL, Dhillon JK, Hammill BG, Dewitt EM, Weinfurt KP, et al. Use of patient-reported outcomes in randomized, double-blind, placebo-controlled clinical trials. *Med Care*. 2011 Apr;49(4):415-9.
80. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes*. 2006;4:79.
81. Greenhalgh J. The applications of PROs in clinical practice: what are they, do they work, and why? *Qual Life Res*. 2009 Feb;18(1):115-23.
82. Santana MJ, Feeny D. Framework to assess the effects of using patient-reported outcome measures in chronic care management. *Qual Life Res*. 2013 Dec 7.
83. Valderas JM, Kotzeva A, Espallargues M, Guyatt G, Ferrans CE, Halyard MY, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Qual Life Res*. 2008 Mar;17(2):179-93.
84. Bausewein C, Simon ST, Benalia H, Downing J, Mwangi-Powell FN, Daveson BA, et al. Implementing patient reported outcome measures (PROMs) in palliative care--users' cry for help. *Health Qual Life Outcomes*. 2011;9:27.
85. Till JE, Osoba D, Pater JL, Young JR. Research on health-related quality of life: dissemination into practical applications. *Qual Life Res*. 1994 Aug;3(4):279-83.
86. Vallance-Owen A, Cubbin S, Warren V, Matthews B. Outcome monitoring to facilitate clinical governance; experience from a national programme in the independent sector. *J Public Health (Oxf)*. 2004 Jun;26(2):187-92.
87. Health and Social Care Information Centre. *Patient Reported Outcome Measures*. UK 2013; Available from: <http://www.hscic.gov.uk/proms>.
88. Smith S, Cano S, Lamping D, Staniszewska S, Browne J, Lewsey J, et al. *Patient-Reported Outcome Measures (PROMs) for routine use in Treatment Centres: recommendations based on a review of the scientific evidence*. London: Department of Health 2005.

89. Lord Darzi. Our NHS Our Future - NHS Next Stage Review Interim Report. London: Department of Health 2007.
90. Department of Health. Guidance on the Standard NHS Contract for Acute Hospital Services. UK 2007 [18/05/14]; Available from: http://www.dhcarenetworks.org.uk/library/guidance_on_the_standard_nhs_contract.pdf.
91. Darzi L. High quality care for all: NHS Next Stage Review final report. London: Department of Health 2008.
92. Department of Health. Guidance on the routine collection of Patient Reported Outcome Measures (PROMs). London 2009.
93. Department of Health. Equity and excellence: Liberating the NHS London 2010.
94. Department of Health. Transparency in outcomes: a framework for the NHS. London 2010.
95. Department of Health. The NHS Outcomes Framework 2011/12. London 2010.
96. Department of Health. The NHS Outcomes Framework 2012/13. London 2011.
97. Department of Health. The NHS Outcomes Framework 2013/14. London 2012.
98. National Health Service. 2014/15 National Tariff Payment System: A Consultation Notice. England 2013.
99. Department of Health. The NHS Outcomes Framework 2014/15. London 2013.
100. Berwick DM, James B, Coye MJ. Connections between quality measurement and improvement. *Med Care*. 2003 Jan;41(1 Suppl):I30-8.
101. Leonard KL, Masatu MC. Using the Hawthorne effect to examine the gap between a doctor's best possible practice and actual performance. *Journal of Development Economics*. [doi: DOI: 10.1016/j.jdeveco.2009.11.001]. November 2010;92(2):226–34.
102. Moriarty JP, Smallman C. En route to a theory of benchmarking. *Benchmarking: An International Journal* 2009;16(4):484-503.
103. Health Services Advisory Group. Medicare Health Outcomes Survey USA 2011 [17/11/11]; Available from: <http://www.hosonline.org/>.
104. 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.
105. Callaly T, Hyland M, Coombs T, Trauer T. Routine outcome measurement in public mental health: results of a clinician survey. *Aust Health Rev*. 2006 May;30(2):164-73.
106. Meehan T, McCombes S, Hatzipetrou L, Catchpole R. Introduction of routine outcome measures: staff reactions and issues for consideration. *J Psychiatr Ment Health Nurs*. 2006 Oct;13(5):581-7.
107. Pirkis J, Burgess P, Coombs T, Clarke A, Jones-Ellis D, Dickson R. Routine measurement of outcomes in Australia's public sector mental health services. *Australia and New Zealand health policy*. 2005 Apr 19;2(1):8.
108. Delnoij DM, Westert GP. Assessing the validity of quality indicators: keep the context in mind! *Eur J Public Health*. 2012 Aug;22(4):452-3.

109. Espallargues M, Valderas JM, Alonso J. Provision of feedback on perceived health status to health care professionals: a systematic review of its impact. *Med Care*. 2000 Feb;38(2):175-86.
110. Gilbody SM, House AO, Sheldon TA. Routinely administered questionnaires for depression and anxiety: systematic review. *BMJ*. 2001 Feb 17;322(7283):406-9.
111. Gilbody SM, House AO, Sheldon T. Routine administration of Health Related Quality of Life (HRQoL) and needs assessment instruments to improve psychological outcome--a systematic review. *Psychol Med*. 2002 Nov;32(8):1345-56.
112. Lockett T, Butow PN, King MT. Improving patient outcomes through the routine use of patient-reported data in cancer clinics: future directions. *Psychooncology*. 2009 Nov;18(11):1129-38.
113. Sutherland K, Coyle N. Quality in Healthcare in England, Wales, Scotland, Northern Ireland: an intra-UK chartbook. London2009.
114. Oduwole KO, Codd MB, Byrne F, O'Byrne J, Kenny PJ. Irish National Joint Registry: a concept. *Ir J Med Sci*. 2008 Dec;177(4):347-53.
115. Okoro T, Ramavath A, Howarth J, Jenkinson J, Maddison P, Andrew JG, et al. What does standard rehabilitation practice after total hip replacement in the UK entail? Results of a mixed methods study. *BMC Musculoskelet Disord*. 2013;14:91.
116. Browne J, Jamieson L, Lewsey J, van der Meulen J, Copley L, Black N. Case-mix & patients' reports of outcome in Independent Sector Treatment Centres: Comparison with NHS providers. *BMC Health Serv Res*. 2008;8:78.
117. Tibrewal S, Sabah S, Henckel J, Hart A. The effect of a manufacturer recall on the threshold to revise a metal-on-metal hip. *International orthopaedics*. 2014 May 15.
118. Economic and Social Research Institute. Activity in Acute Public Hospitals in Ireland. Dublin2011.
119. Economic and Social Research Institute. Activity in Acute Public Hospitals in Ireland. Dublin2012.
120. National Office of Clinical Audit. NOCA Audits - Irish National Orthopaedic Register (INOR). Ireland2014 [18/05/14]; Available from: <http://www.noca.ie/noca-inor>.
121. Health Service Executive. Health Service National Performance Assurance Report. Ireland: Health Service Executive2013 AUGUST 2013.
122. Department of Health. Future Health- A Strategic Framework for Reform of the Health Service 2012 – 2015. Dublin2012.
123. Harvey B. Evolution of health services and health policy in ireland. Dublin: Combat Poverty Agency2007.
124. O'Ferrall. Universal Health Insurance-What is it and would it be effective in Ireland? Working Notes. 2009(60).
125. Burke S, editor. Irish Apartheid, Healthcare Inequality in Ireland. Dublin: New Island; 2009.
126. Higgins J. The Establishment of Hospital Groups as a transition to Independent Hospital Trusts- A report to the Minister for Health, Dr James Reilly, TD. Dublin: Department of Health2012 February 2013.
127. Department of Health. Primary Care- A New Direction. Quality and Fairness - A Health System for You. Health Strategy. Dublin2001.
128. McDaid D, Wiley M, Maresso A, Mossialos E. Health Systems in Transition. Ireland- Health system review. Copenhagen: WHO2009.

129. Harding Clark M. The Lourdes Hospital Inquiry- An Inquiry into peripartum hysterectomy at Our Lady of Lourdes Hospital, Drogheda. Dublin: Government of Ireland 2006.
130. Health Information and Quality Authority. Report of the investigation into the quality, safety and governance of the care provided by the Adelaide and Meath Hospital, Dublin incorporating the National Children's Hospital (AMNCH) for patients who require acute admission. Ireland 2012 8 May 2012.
131. Health Information and Quality Authority. National Quality Assurance Standards for Symptomatic Breast Disease Services. Ireland 2007 May 18, 2007.
132. Health Service Executive. Draft National Clinical Audit Guidance Document. Ireland: National Director of Quality and Patient Safety 2012.
133. Health Service Executive. Medical Consultants' Contract - Comptroller and Auditor General Special Report. Ireland 2007.
134. Medical Council. Professional Competence- Guidelines for Doctors. Ireland 2012.
135. Department of Health. Special Deliveries Unit. Ireland 2014 [18/05/14]; Available from: http://www.dohc.ie/about_us/divisions/special_delivery_unit.
136. Special Deliveries Unit. Special Delivery Unit Unscheduled Care Strategic Plan. Ireland: HSE 2013 Quarter 1, 2013.
137. Royal College of Surgeons Ireland. Welcome to the National Surgical Training Centre. 2014; Available from: http://www.rcsi.ie/surgery_nstc.
138. Sayana MK, Ashraf M, O'Byrne J. Modernising the higher surgical training in trauma and orthopaedic surgery in Ireland: taking the middle path approach. *Ir J Med Sci*. 2009 Dec;178(4):389-92.
139. Creswell JW, Plano Clark VL. Designing and conducting mixed methods research. USA: Sage publications Inc.; 2011.
140. Creswell JW. Research Design. Qualitative, Quantitative and mixed methods approaches. UK: Sage publications Inc.; 2014.
141. 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 Mar 17.
142. Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. *BMJ Qual Saf*. 2014 Feb 6.
143. Boyce MB, Browne JP, Greenhalgh J. Surgeon's experiences of receiving peer benchmarked feedback using patient-reported outcome measures: a qualitative study. *Implementation Science: IS*. 2014.
144. Rosenbloom SK, Victorson DE, Hahn EA, Peterman AH, Cella D. Assessment is not enough: a randomized controlled trial of the effects of HRQL assessment on quality of life and satisfaction in oncology clinical practice. *Psychooncology*. 2007 Dec;16(12):1069-79.
145. Greenhalgh J, Meadows K. The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. *J Eval Clin Pract*. 1999 Nov;5(4):401-16.
146. Deyo RA, Carter WB. Strategies for improving and expanding the application of health status measures in clinical settings. A researcher-developer viewpoint. *Med Care*. 1992 May;30(5 Suppl):MS176-86; discussion MS96-209.
147. 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 Feb;60(4):833-43.

148. Cano S, Hobart J. The problem with health measurement. *Patient Preference And Adherence*. 2011;5:279-90.
149. Sutherland HJ, Till JE. Quality of life assessments and levels of decision making: differentiating objectives. *Qual Life Res*. 1993 Aug;2(4):297-303.
150. Keogh B, Spiegelhalter D, Bailey A, Roxburgh J, Magee P, Hilton C. The legacy of Bristol: public disclosure of individual surgeons' results. *BMJ*. 2004 Aug 21;329(7463):450-4.
151. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? *Ann Intern Med*. 2006 Aug 15;145(4):265-72.
152. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*. 2006(2):CD000259.
153. Fung CH, Lim YW, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. *Ann Intern Med*. 2008 Jan 15;148(2):111-23.
154. Coulter A, Ellins J. Patient-focused interventions. A review of the evidence: The Health Foundation 2006.
155. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. *J Eval Clin Pract*. 2006 Oct;12(5):559-68.
156. Department of Health & Children. Primary Care: A New Direction. Dublin: Government of Ireland 2001.
157. Expert Advisory Group on Cancer. A policy framework for commissioning cancer services: a report to the chief medical officers of England and Wales. The Calman-Hine Report. London 1995.
158. Mental Health Commission. Multidisciplinary Team Working: From Theory to Practice- Discussion Paper. Dublin 2006.
159. Robinson KA, Dickersin K. Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. *Int J Epidemiol*. 2002 Feb;31(1):150-3.
160. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*. 1996 Feb;17(1):1-12.
161. Fayers PM. Evaluating the effectiveness of using PROs in clinical practice: a role for cluster-randomised trials. *Qual Life Res*. 2008 Dec;17(10):1315-21.
162. Cohen J. Statistical power analysis for the behavioral sciences. New Jersey: Lawrence Erlbaum Associates Inc. ; 1969.
163. Calkins DR, Rubenstein LV, Cleary PD, Davies AR, Jette AM, Fink A, et al. Functional disability screening of ambulatory patients: a randomized controlled trial in a hospital-based group practice. *J Gen Intern Med*. 1994 Oct;9(10):590-2.
164. Dowrick C, Buchan I. Twelve month outcome of depression in general practice: does detection or disclosure make a difference? *BMJ*. 1995 Nov 11;311(7015):1274-6.
165. German PS, Shapiro S, Skinner EA, Von Korff M, Klein LE, Turner RW, et al. Detection and management of mental health problems of older patients by primary care providers. *JAMA*. 1987 Jan 23-30;257(4):489-93.
166. Kazis LE, Callahan LF, Meenan RF, Pincus T. Health status reports in the care of patients with rheumatoid arthritis. *J Clin Epidemiol*. 1990;43(11):1243-53.

167. Mathias SD, Fifer SK, Mazonson PD, Lubeck DP, Buesching DP, Patrick DL. Necessary but not sufficient: the effect of screening and feedback on outcomes of primary care patients with untreated anxiety. *J Gen Intern Med.* 1994 Nov;9(11):606-15.
168. Rubenstein LV, Calkins DR, Young RT, Cleary PD, Fink A, Kosecoff J, et al. Improving patient function: a randomized trial of functional disability screening. *Ann Intern Med.* 1989 Nov 15;111(10):836-42.
169. Rubenstein LV, McCoy JM, Cope DW, Barrett PA, Hirsch SH, Messer KS, et al. Improving patient quality of life with feedback to physicians about functional status. *J Gen Intern Med.* 1995 Nov;10(11):607-14.
170. White P, Atherton A, Hewett G, Howells K. Using information from asthma patients: a trial of information feedback in primary care. *BMJ.* 1995 Oct 21;311(7012):1065-9.
171. Whooley MA, Stone B, Soghikian K. Randomized trial of case-finding for depression in elderly primary care patients. *J Gen Intern Med.* 2000 May;15(5):293-300.
172. Wasson JH, Stukel TA, Weiss JE, Hays RD, Jette AM, Nelson EC. A randomized trial of the use of patient self-assessment data to improve community practices. *Eff Clin Pract.* 1999 Jan-Feb;2(1):1-10.
173. Trowbridge R, Dugan W, Jay SJ, Littrell D, Casebeer LL, Edgerton S, et al. Determining the effectiveness of a clinical-practice intervention in improving the control of pain in outpatients with cancer. *Acad Med.* 1997 Sep;72(9):798-800.
174. Lambert MJ, Whipple JL, Smart DW, Vermeersch DA, Nielsen SL. The Effects of Providing Therapists With Feedback on Patient Progress During Psychotherapy: Are Outcomes Enhanced? *Psychotherapy Research* 2001;11(1):49-68.
175. Hawkins EJ, Lambert MJ, Vermeersch DA, Slade KL, Tuttle KC. The therapeutic effects of providing patient progress information to therapists and patients. *Psychotherapy Research.* 2004;14(3):308-27.
176. McCoy JM, Rubenstein L, Hirsch SH, Barrett PA. A Feedforward System for Functional Status Information. *Proc Annu Symp Comput Appl Med Care.* 1988;9:683-6. .
177. Puschner B, Schöfer D, Knaup C, Becker T. Outcome management in in-patient psychiatric care. *Acta Psychiatr Scand* 2009;120(4):308-19.
178. Gutteling JJ, Darlington AS, Janssen HL, Duivenvoorden HJ, Busschbach JJ, de Man RA. Effectiveness of health-related quality-of-life measurement in clinical practice: a prospective, randomized controlled trial in patients with chronic liver disease and their physicians. *Qual Life Res.* 2008 Mar;17(2):195-205.
179. Weingarten SR, Kim CS, Stone EG, Kristopaitis RJ, Pelter M, Sandhu M. Can peer-comparison feedback improve patient functional status? *Am J Manag Care.* 2000 Jan;6(1):35-9.
180. Berking M, Orth U, Lutz W. How effective is systematic feedback of treatment progress to the therapist? An empirical study in a cognitive-behavioural-oriented inpatient setting. *Zeitschrift für Klinische Psychologie und Psychotherapie.* 2006;35:21-9.
181. Arah OA, Klazinga NS, Delnoij DM, ten Asbroek AH, Custers T. Conceptual frameworks for health systems performance: a quest for effectiveness, quality, and improvement. *Int J Qual Health Care.* 2003 Oct;15(5):377-98.
182. Lea W. The NHS Outcomes Framework 2014/15: Department of Health 2013.

183. Advancing Quality Alliance. Advancing Quality 3 Year Business Plan 2013/14 – 2015/16. Salford: Advancing Quality Programme Office 2013.
184. Department of Health. Analysis of Patient Reported Outcomes Measures. Pre-Operative Health Status Data- April 2009 to March 2010
England: Department of Health 2011.
185. National Health Service. Patient Reported Outcome Measures- PROMs statistics NHS Choices; 2013 [03/01/14]; Available from:
<http://www.nhs.uk/NHSEngland/thenhs/records/proms/Pages/statistics.aspx>.
186. Gawande A, editor. Better- A surgeon's notes on performance. London: Profile Books Ltd; 2007.
187. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. *J Clin Oncol*. 2004 Feb 15;22(4):714-24.
188. Soreide K, Soreide AH. Using patient-reported outcome measures for improved decision-making in patients with gastrointestinal cancer - the last clinical frontier in surgical oncology? *Frontiers in oncology*. 2013;3:157.
189. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. *BMC Health Serv Res*. 2013;13:211.
190. Schunemann HJ, Akl EA, Guyatt GH. Interpreting the results of patient reported outcome measures in clinical trials: the clinician's perspective. *Health Qual Life Outcomes*. 2006;4:62.
191. Trauer T, Gill L, Pedwell G, Slattery P. Routine outcome measurement in public mental health--what do clinicians think? *Aust Health Rev*. 2006 May;30(2):144-7.
192. Duncan EA, Murray J. The barriers and facilitators to routine outcome measurement by allied health professionals in practice: a systematic review. *BMC Health Serv Res*. 2012;12:96.
193. Brettell AJ, Long AF, Grant MJ, Greenhalgh J. Searching for information on outcomes: do you need to be comprehensive? *Qual Health Care*. 1998 Sep;7(3):163-7.
194. McKibbin KA, Wilczynski NL, Haynes RB. Developing optimal search strategies for retrieving qualitative studies in PsycINFO. *Eval Health Prof*. 2006 Dec;29(4):440-54.
195. Critical Appraisal Skills Programme. Making sense of evidence about clinical effectiveness. 10 questions to help you make sense of qualitative research. UK2013; Qualitative checklist_14.10.10:[Available from: <http://www.casp-uk.net/>].
196. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med Res Methodol*. 2008;8:45.
197. QSR International. NVivo qualitative data analysis software Version 10 ed 2013.
198. Cranley L, Doran DM. Nurses' integration of outcomes assessment data into practice. *Outcomes Manag*. 2004 Jan-Mar;8(1):13-8.
199. Dowrick C, Leydon GM, McBride A, Howe A, Burgess H, Clarke P, et al. Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. *BMJ*. 2009;338:b663.
200. Dunckley M, Aspinall F, Addington-Hall JM, Hughes R, Higginson IJ. A research study to identify facilitators and barriers to outcome measure implementation. *Int J Palliat Nurs*. 2005 May;11(5):218-25.

201. Eischens MJ, Elliott BA, Elliott TE. Two hospice quality of life surveys: a comparison. *Am J Hosp Palliat Care*. 1998 May-Jun;15(3):143-8.
202. Hughes R, Aspinall F, Addington-Hall J, Chidgey J, Drescher U, Dunckley M, et al. Professionals' views and experiences of using outcome measures in palliative care. *Int J Palliat Nurs*. 2003 Jun;9(6):234-8.
203. Hughes R, Aspinall F, Addington-Hall JM, Dunckley M, Faull C, Higginson I. It just didn't work: the realities of quality assessment in the English health care context. *International Journal of Nursing Studies*. [doi: 10.1016/j.ijnurstu.2004.02.005]. 2004;41(7):705-12.
204. Kettis-Lindblad A, Ring L, Widmark E, Bendtsen P, Glimelius B. Patients' and doctors' views of using the schedule for individual quality of life in clinical practice. *J Support Oncol*. 2007 Jun;5(6):281-7.
205. Mason L, Poole H. Healthcare professionals' views of screening for postnatal depression. *Community Pract*. 2008 Apr;81(4):30-3.
206. Mitchell C, Dwyer R, Hagan T, Mathers N. Impact of the QOF and the NICE guideline in the diagnosis and management of depression: a qualitative study. *Br J Gen Pract*. 2011 May;61(586):e279-89.
207. Slater A, Freeman E. Is the Palliative Care Outcome Scale useful to staff in a day hospice unit? *International Journal Of Palliative Nursing*. 2005;11(7):346-54.
208. Tavabie JA, Tavabie OD. Improving care in depression: qualitative study investigating the effects of using a mental health questionnaire. *Qual Prim Care*. 2009;17(4):251-61.
209. Unsworth G, Cowie H, Green A. Therapists' and clients' perceptions of routine outcome measurement in the NHS: A qualitative study. *Counselling & Psychotherapy Research*. 2012;12(1):71-80.
210. Wressle EJMJC. The Canadian Occupational Performance Measure as an outcome measure and team tool in a day treatment programme. *Disability & Rehabilitation*. [Article]. 2003 05/20;25(10):497.
211. Sammer CE, Lykens K, Singh KP, Mains DA, Lackan NA. What is patient safety culture? A review of the literature. *J Nurs Scholarsh*. 2010 Jun;42(2):156-65.
212. Close-Goedjen JL, Saunders SM. The effect of technical support on clinician attitudes toward an outcome assessment instrument. *The Journal Of Behavioral Health Services & Research*. 2002;29(1):99-108.
213. Coleman NE, Pon S. Quality: performance improvement, teamwork, information technology and protocols. *Critical care clinics*. 2013 Apr;29(2):129-51.
214. Lemire M, Demers-Payette O, Jefferson-Falardeau J. Dissemination of performance information and continuous improvement: A narrative systematic review. *J Health Organ Manag*. 2013;27(4):449-78.
215. Kaplan HC, Brady PW, Dritz MC, Hooper DK, Linam WM, Froehle CM, et al. The influence of context on quality improvement success in health care: a systematic review of the literature. *The Milbank quarterly*. 2010 Dec;88(4):500-59.
216. Wilkinson J, Powell A, Davies H. Are clinicians engaged in quality improvement? A review of the literature on healthcare professionals' views on quality improvement initiatives. London: The Health Foundation 2011 May 2011.
217. Moran P, Kelesidi K, Guglani S, Davidson S, Ford T. What do parents and carers think about routine outcome measures and their use? A focus group study of CAMHS attenders. *Clin Child Psychol Psychiatry*. 2012 Jan;17(1):65-79.
218. Matata B, Hinder S, Steele S, Gibbons E, Jackson M. Patients' attitudes and perceptions of two health-related quality-of-life questionnaires used to collect patient-reported outcome measures in the English National Health Service: A

qualitative study of patients undergoing cardiac interventions. *SAGE Open Medicine* 2013;1(1).

219. Hysong SJ, Best RG, Pugh JA. Audit and feedback and clinical practice guideline adherence: making feedback actionable. *Implement Sci.* 2006;1:9.
220. Wood-Dauphinee S. Assessing quality of life in clinical research: from where have we come and where are we going? *J Clin Epidemiol.* 1999 Apr;52(4):355-63.
221. Bickman L, Kelley SD, Breda C, de Andrade AR, Riemer M. Effects of routine feedback to clinicians on mental health outcomes of youths: results of a randomized trial. *Psychiatr Serv.* 2011 Dec;62(12):1423-9.
222. Bovey WH, Hede A. Resistance to organisational change: the role of defence mechanisms. *Journal of Managerial Psychology.* 2001;16 (7):534 - 48.
223. Hobart J, Cano S. Rating Scales for Clinical Studies in Neurology—Challenges and Opportunities. *Usneurology.* 2008;4(1):12-8.
224. Wolpert M. Uses and Abuses of Patient Reported Outcome Measures (PROMs): Potential Iatrogenic Impact of PROMs Implementation and How It Can Be Mitigated. *Adm Policy Ment Health.* 2013 Jul 19.
225. Wolpert M. Do patient reported outcome measures do more harm than good? *BMJ.* 2013;346:f2669.
226. 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 Nov;22(9):2265-78.
227. Varagunam M, Hutchings A, Neuburger J, Black N. Impact on hospital performance of introducing routine patient reported outcome measures in surgery. *J Health Serv Res Policy.* 2013 Sep 26.
228. National Quality Forum. Patient Reported Outcomes (PROs) in Performance Measurement. Washington: National Quality Forum 2013 January 10, 2013.
229. MRC. Cluster randomised trials: Methodological and ethical considerations: Medical Research Council 2002.
230. Hildon Z, Neuburger J, Allwood D, van der Meulen J, Black N. 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.
231. Dawson J, Fitzpatrick R, Churchman D, Verjee-Lorenz A, Clayson D. User Manual for the Oxford Hip Score (OHS). Information on development and use of the OHS. Oxford: University of Oxford 2010.
232. Department of Health. Patient Reported Outcome Measures (PROMs) in England. The case-mix adjustment methodology 2012 11 April 2012.
233. Davis AM, Perruccio AV, Canizares M, Hawker GA, Roos EM, Maillefert JF, 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 Jul;17(7):843-7.
234. Cheung K, Oemar M, Oppe M, Rabin R. User Guide- basic information on how to use EQ5D: EuroQol Group 2009.
235. Norman GR, Sloan JA, Wywich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care.* 2003 May;41(5):582-92.
236. Altman DG. Randomisation. *BMJ.* 1991 Jun 22;302(6791):1481-2.
237. Coleman K, Hamblin R. Can pay-for-performance improve quality and reduce health disparities? *PLoS Med.* 2007 Jun;4(6):e216.

238. Deutsch A, Smith L, Gage B, Kelleher C, Garfinkel D. Patient-Reported Outcomes in Performance Measurement: National Quality Forum 2012.
239. Lyratzopoulos G, Elliott MN, Barbiere JM, Staetsky L, Paddison CA, Campbell J, et al. How can Health Care Organizations be Reliably Compared?: Lessons From a National Survey of Patient Experience. *Med Care*. 2011 Aug;49(8):724-33.
240. Singh JA, Kwok CK, Boudreau RM, Lee GC, Ibrahim SA. Hospital volume and surgical outcomes after elective hip/knee arthroplasty: a risk-adjusted analysis of a large regional database. *Arthritis and rheumatism*. 2011 Aug;63(8):2531-9.
241. Lingard L, editor. Power of the NHS Patient Reported Outcome Measures (PROMs) Programme to Improve Quality. ISQua; 2013; Edinburgh: ISQua.
242. Jahagirdar D, Kroll T, Ritchie K, Wyke S. 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.
243. Fung V, Schmittiel JA, Fireman B, Meer A, Thomas S, Smider N, et al. Meaningful variation in performance: a systematic literature review. *Med Care*. 2010 Feb;48(2):140-8.
244. Hannan EL, Kilburn H, Jr., Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA*. 1994 Mar 9;271(10):761-6.
245. Slade M. The use of patient-level outcomes to inform treatment. *Epidemiol Psychiatr Soc*. 2002 Jan-Mar;11(1):20-7.
246. Bendtsen P, Leijon M, Sofie Sommer A, Kristenson M. Measuring health-related quality of life in patients with chronic obstructive pulmonary disease in a routine hospital setting: feasibility and perceived value. *Health And Quality Of Life Outcomes*. 2003;1:5-.
247. Appleby J. PROMs: Counting what matters most to patients. *The Kings Fund*; 2009 [11/05/2010]; Available from: www.kingsfund.org.uk/blog/proms_counting_what.html.
248. National Health Service. Equity and excellence: Liberating the NHS. London: Department of Health 2010.
249. Valderas JM, Fitzpatrick R, Roland M. Using health status to measure NHS performance: another step into the dark for the health reform in England. *BMJ Qual Saf*. 2012 Apr;21(4):352-3.
250. National Health Service. Patient Reported Outcome Measures (PROMs). An alternative aggregation methodology for case-mix adjustment. London 2013.
251. EuroQol. About EQ5D. The Netherlands: EuroQol Group Executive Office; 2014; Available from: <http://www.euroqol.org/about-eq-5d.html>.
252. Davis AM, Perruccio AV, Canizares M, Tennant A, Hawker GA, Conaghan PG, et al. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. *Osteoarthritis Cartilage*. 2008 May;16(5):551-9.
253. Richie J, Lewis J. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*- Chapter 9 "Carrying out Qualitative Analysis. London: Sage Publications Ltd.; 2003.
254. Green J, Thorogood N. *Qualitative Methods for Health Researchers*. London: Sage Publications Ltd; 2004.
255. National Centre for Social Research. Framework: The Framework method for qualitative data analysis. 2013; Available from: <http://www.natcen.ac.uk/our-expertise/framework>.

256. University College Cork. An Introduction to Research Ethics at UCC. Cork: University College Cork 2007.
257. Ziebland S, McPherson A. Making sense of qualitative data analysis: an introduction with illustrations from DIPEX (personal experiences of health and illness). *Med Educ.* 2006 May;40(5):405-14.
258. Baker SE, Edwards R. How many qualitative interviews is enough. . 2012.
259. Harvey WS. Methodological Approaches for Junior Researchers Interviewing Elites: A Multidisciplinary Perspective. Economic Geography Research Group. [Working Paper Series]. In press.
260. Francis JJ, Johnston M, Robertson C, Glidewell L, Entwistle V, Eccles MP, et al. What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology & health.* 2010 Dec;25(10):1229-45.
261. Shenton AK. Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information* 2004 22 63–75.
262. Alazzawi S, Bardakos NV, Hadfield SG, Butt U, Beer ZH, Field RE. Patient-reported complications after elective joint replacement surgery: are they correct? *J Bone Joint Surg Br.* 2012 Aug;94(8):1120-5.
263. Berkanovic E, Hurwicz ML, Lachenbruch PA. Concordant and discrepant views of patients' physical functioning. *Arthritis Care Res.* 1995 Jun;8(2):94-101.
264. Detmar SB, Aaronson NK, Wever LD, Muller M, Schornagel JH. How are you feeling? Who wants to know? Patients' and oncologists' preferences for discussing health-related quality-of-life issues. *J Clin Oncol.* 2000 Sep 15;18(18):3295-301.
265. Middleton S, Lumby J. Comparing professional and patient outcomes for the same episode of care. *Aust J Adv Nurs.* 1999 Sep-Nov;17(1):22-7.
266. Newell S, Sanson-Fisher RW, Girgis A, Bonaventura A. How well do medical oncologists' perceptions reflect their patients' reported physical and psychosocial problems? Data from a survey of five oncologists. *Cancer.* 1998 Oct 15;83(8):1640-51.
267. Suarez-Almazor ME, Conner-Spady B, Kendall CJ, Russell AS, Skeith K. Lack of congruence in the ratings of patients' health status by patients and their physicians. *Medical Decision Making: An International Journal Of The Society For Medical Decision Making.* 2001;21(2):113-21.
268. Allwood D, Hildon Z, Black N. Clinicians' views of formats of performance comparisons. *Journal of Evaluation in Clinical Practice.* 2013;19(1):86-93.
269. Hildon Z, Neuburger J, Allwood D, van der Meulen J, Black N. Clinicians' and patients' views of metrics of change derived from patient reported outcome measures (PROMs) for comparing providers' performance of surgery. *BMC Health Services Research.* 2012;12:171-.
270. Cano SJ, Barrett LE, Zajicek JP, Hobart JC. Beyond the reach of traditional analyses: using Rasch to evaluate the DASH in people with multiple sclerosis. *Multiple sclerosis (Houndmills, Basingstoke, England).* 2011 Feb;17(2):214-22.
271. Bamm EL, Rosenbaum P, Wilkins S. Is Health Related Quality Of Life of people living with chronic conditions related to patient satisfaction with care? *Disability & Rehabilitation.* 2013;35(9):766-74.
272. Awad SS. Adherence to surgical care improvement project measures and post-operative surgical site infections. *Surg Infect (Larchmt).* 2012 Aug;13(4):234-7.
273. Stulberg JJ, Delaney CP, Neuhauser DV, Aron DC, Fu P, Koroukian SM. Adherence to surgical care improvement project measures and the association with postoperative infections. *JAMA.* 2010 Jun 23;303(24):2479-85.

274. Hall BL, Hamilton BH, Richards K, Bilimoria KY, Cohen ME, Ko CY. Does surgical quality improve in the American College of Surgeons National Surgical Quality Improvement Program: an evaluation of all participating hospitals. *Ann Surg*. 2009 Sep;250(3):363-76.
275. Share DA, Campbell DA, Birkmeyer N, Prager RL, Gurm HS, Moscucci M, et al. How a regional collaborative of hospitals and physicians in Michigan cut costs and improved the quality of care. *Health Aff (Millwood)*. 2011 Apr;30(4):636-45.
276. Kotronoulas G, Kearney N, Maguire R, Harrow A, Di Domenico D, Croy S, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. *J Clin Oncol*. 2014 May 10;32(14):1480-501.
277. Callaly T, Hyland M, Coombs T, Trauer T. Routine outcome measurement in public mental health: results of a clinician survey. *Australian Health Review*. 2006;30(2):164-73.
278. Meehan T, McCombes S, Hatzipetrou L, Catchpoole R. Introduction of routine outcome measures: staff reactions and issues for consideration. *Journal Of Psychiatric And Mental Health Nursing*. 2006;13(5):581-7.
279. Kettis-Lindblad A, Ring L, Widmark E, Bendtsen P, Glimelius B. Patients'and doctors' views of using the schedule for individual quality of life in clinical practice. *The Journal Of Supportive Oncology*. 2007;5(6):281-7.
280. Antunes B, Harding R, Higginson IJ. Implementing patient-reported outcome measures in palliative care clinical practice: a systematic review of facilitators and barriers. *Palliat Med*. 2014 Feb;28(2):158-75.
281. Economic Policy Unit. *A Practical Guide to Policy Making in Northern Ireland*. Belfast: Office of the First Minister and Deputy First Minister.
282. Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data: what do we expect to gain? A review of the evidence. *JAMA*. 2000 Apr 12;283(14):1866-74.
283. Walker IA, Reshamwalla S, Wilson IH. Surgical safety checklists: do they improve outcomes? *Br J Anaesth*. 2012 Jul;109(1):47-54.
284. de Vries EN, Prins HA, Crolla RM, den Outer AJ, van An del G, van Helden SH, et al. Effect of a comprehensive surgical safety system on patient outcomes. *N Engl J Med*. 2010 Nov 11;363(20):1928-37.
285. Cataife G, Weinberg DA, Wong HH, Kahn KL. The effect of Surgical Care Improvement Project (SCIP) compliance on surgical site infections (SSI). *Med Care*. 2014 Feb;52(2 Suppl 1):S66-73.
286. Skeat J, Perry A. Exploring the implementation and use of outcome measurement in practice: a qualitative study. *Int J Lang Commun Disord*. 2008 Mar-Apr;43(2):110-25.
287. Petticrew M. 'More research needed': plugging gaps in the evidence base on health inequalities. *Eur J Public Health*. 2007 Oct;17(5):411-3.
288. Loeb JM. The current state of performance measurement in health care. *Int J Qual Health Care*. 2004 Apr;16 Suppl 1:i5-9.
289. *Functionality and feedback: A realist synthesis of the collation, interpretation and utilisation of PROMs*, (2012).
290. Oxman AD, Fretheim A, Flottorp S. The OFF theory of research utilization. *J Clin Epidemiol*. 2005 Feb;58(2):113-6; discussion 7-20.

291. Balas EA, Boren SA, Brown GD, Ewigman BG, Mitchell JA, Perkoff GT. Effect of physician profiling on utilization. Meta-analysis of randomized clinical trials. *J Gen Intern Med.* 1996 Oct;11(10):584-90.
292. Shekelle P, Lim Y, Mattke S, Damberg C. Does public release of performance results improve quality of care? A systematic review: The Health Foundation 2008.
293. McClimans LM, Browne J. Choosing a patient-reported outcome measure. *Theor Med Bioeth.* 2011 Feb;32(1):47-60.
294. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010 Jul;63(7):737-45.
295. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg Br.* 1998 Jan;80(1):63-9.
296. Wylde V, Learmonth ID, Cavendish VJ. The Oxford hip score: the patient's perspective. *Health Qual Life Outcomes.* 2005;3:66.
297. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br.* 1996 Mar;78(2):185-90.
298. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess.* 2009 Feb;13(12):iii, ix-x, 1-177.
299. Massof RW. Understanding Rasch and item response theory models: applications to the estimation and validation of interval latent trait measures from responses to rating scale questionnaires. *Ophthalmic epidemiology.* 2011 Feb;18(1):1-19.
300. Streiner DL, Norman GR, editors. *Health Measurement Scales—a practical guide to their development and use.* Fourth Edition. New York: Oxford University Press; 2008.
301. Audit Commission. *Improving information to support decision making: standards for better quality data.* England: NHS 2007.
302. Wait S, Nolte E. Benchmarking health systems: trend, conceptual issues and future perspectives. *Benchmarking: An International Journal.* 2005;12(5):436-48.
303. Hjollund NH, Larsen LP, Biering K, Johnsen SP, Riiskjaer E, Schougaard LM. Use of Patient-Reported Outcome (PRO) Measures at Group and Patient Levels: Experiences From the Generic Integrated PRO System, WestChronic. *Interactive journal of medical research.* 2014;3(1):e5.
304. Zbrozek A, Hebert J, Gogates G, Thorell R, Dell C, Molsen E, et al. Validation of electronic systems to collect patient-reported outcome (PRO) data—recommendations for clinical trial teams: report of the ISPOR ePRO systems validation good research practices task force. *Value Health.* 2013 Jun;16(4):480-9.
305. Jahagirdar D, Kroll T, Ritchie K, Wyke S. Patient-reported outcome measures for chronic obstructive pulmonary disease: the exclusion of people with low literacy skills and learning disabilities. *The patient.* 2013;6(1):11-21.
306. Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. *BMJ.* 2010;340:c186.
307. Ganz PA. Quality of Life and the Patient with Cancer. Individual and Policy Implications. *Cancer supplement.* 1994 August 15;74(4):1445-52.

308. Hutchings A, Grosse Frie K, Neuburger J, van der Meulen J, Black N. Late response to patient-reported outcome questionnaires after surgery was associated with worse outcome. *J Clin Epidemiol*. 2013 Feb;66(2):218-25.
309. Hutchings A, Neuburger J, Grosse Frie K, Black N, van der Meulen J. Factors associated with non-response in routine use of patient reported outcome measures after elective surgery in England. *Health Qual Life Outcomes*. 2012;10:34.
310. de Jonge V, Sint Nicolaas J, van Leerdam ME, Kuipers EJ. Overview of the quality assurance movement in health care. *Best Pract Res Clin Gastroenterol*. 2011 Jun;25(3):337-47.
311. Brozek JL, Guyatt GH, Schunemann HJ. How a well-grounded minimal important difference can enhance transparency of labelling claims and improve interpretation of a patient reported outcome measure. *Health Qual Life Outcomes*. 2006;4:69.
312. Kennerley M, Mason S. The Use of Information in Decision Making- Literature Review for the Audit Commission: Centre for Business Performance 2008.
313. Lohr KN, Aaronson NK, Alonso J, Audrey Burnam M, Patrick DL, Perrin EB, et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clinical Therapeutics*. [doi: DOI: 10.1016/S0149-2918(96)80054-3]. 1996;18(5):979-92.
314. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc*. 2002 Apr;77(4):371-83.
315. Tanenbaum SJ. Evidence and expertise: the challenge of the outcomes movement to medical professionalism. *Acad Med*. 1999 Jul;74(7):757-63.
316. Lyratzopoulos G, editor. How can healthcare providers be reliably compared? Lessons from a national survey of patient experience. Delivering better health services - Health Services Research Network; 2011 07/06/11; ACC Liverpool. <http://www.nhsconfed.org/Events/events-archive/deliveringbetterhealthservices/Pages/Sessions-Dayone.aspx2011>.
317. Hahn EA, Cella D, Chassany O, Fairclough DL, Wong GY, Hays RD. Precision of health-related quality-of-life data compared with other clinical measures. *Mayo Clin Proc*. 2007 Oct;82(10):1244-54.
318. Browne JP, Bastaki H, Dawson J. What is the optimal time point to assess patient-reported recovery after hip and knee replacement? A systematic review and analysis of routinely reported outcome data from the English patient-reported outcome measures programme. *Health Qual Life Outcomes*. 2013;11:128.
319. Doran T, Kontopantelis E, Valderas JM, Campbell S, Roland M, Salisbury C, et al. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. *BMJ*. 2011;342:d3590.
320. Smith PC, Mossialos E, Papanicolas I. Performance measurement for health system improvement: experiences, challenges and prospects. Copenhagen: World Health Organisation 2008.
321. Reeves D, Doran T, Valderas JM, Kontopantelis E, Trueman P, Sutton M, et al. How to identify when a performance indicator has run its course. *BMJ*. 2010;340:c1717.
322. Addington D, Kyle T, Desai S, Wang J. Facilitators and barriers to implementing quality measurement in primary mental health care: Systematic review. *Can Fam Physician*. 2010 Dec;56(12):1322-31.

323. Alonso J, Bartlett SJ, Rose M, Aaronson NK, Chaplin JE, Efficace F, et al. The case for an international patient-reported outcomes measurement information system (PROMIS(R)) initiative. *Health Qual Life Outcomes*. 2013;11:210.
324. Albert W. *Advances in the Use of Patient Reported Outcome Measures in Electronic Health Records*. Baltimore, Maryland: Center for Health Services and Outcomes Research 2013.
325. Council MR. *Cluster randomised trials: Methodological and ethical considerations* 2002.
326. Eldridge S, Kerry S. *A Practical Guide to Cluster Randomised Trials in Health Services Research*. UK: Wiley; 2012.

Appendices

Appendix 1: Quantitative Search Strategy

1. score*
2. questionnaire*
3. scale*
4. measure*
5. instrument*
6. (#1 OR #2 OR #3 OR #4 OR #5)
7. patient based
8. self report*
9. patient report*
10. patient related
11. patient*
12. (#7 OR #8 OR #9 OR #10 OR #11)
13. performance status
14. disability scale
15. functional status
16. quality of life
17. health status
18. (#13 OR #14 OR #15 OR #16 OR #17)
19. (#6 OR #12 OR #18)
20. Feedback

21. Audit
22. (#20 OR #21)
23. "Physician-Patient Relations"[Mesh]
24. "Clinical Competence"[Mesh])
25. "Physician's Practice Patterns/standards"[Mesh]
26. (#23 OR #24 OR #25)
27. Quasi-randomised trial
28. Non-randomised trial
29. quasi-experimental study
30. Controlled before-and-after study
31. Before-and-after study
32. (#27 OR #28 OR #29 OR #30 OR #31)
33. randomized controlled trial[pt]
34. controlled clinical trial[pt]
35. randomized controlled trials[mh]
36. random allocation[mh]
37. double-blind method[mh]
38. single-blind method[mh]
39. clinical trial[pt]
40. clinical trials[mh]
41. "clinical trial"[tw]
42. (#33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39 OR #40 OR #41)

43. singl*[tw]
44. doubl*[tw]
45. trebl*[tw]
46. tripl*[tw]
47. (#43 OR #44 OR #45 OR #46)
48. mask*[tw]
49. blind*[tw]))
50. (#48 OR #49)
51. (#47 AND #50)
52. ("latin square"[tw])
53. placebos[mh]
54. placebo*[tw]
55. random*[tw]
56. research design[mh:noexp]
57. (#52 OR #53 OR #54 OR #55 OR #56)
58. comparative study[mh]
59. evaluation studies[mh]
60. follow-up studies[mh]
61. prospective studies[mh]
62. cross-over studies[mh]
63. control*[tw]
64. prospectiv*[tw]

65. volunteer*[tw])

66. (#58 OR #59 OR #60 OR #61 OR #62 OR #63 OR #63 OR #64 OR #65)

67. (#32 OR #42 OR #51 OR #57 OR #66)

68. (animal[mh] NOT human[mh]))))

69. (#67 NOT #68)

70. (#19 AND #22 AND #26 AND #69)

Appendix 2: Qualitative Search Strategy

Medline

Block 1

1. (MH "health status indicator")
2. (MH "outcome and process assessment (health care)")
3. (MH "outcome assessment (health care)")
4. (MH "quality of life")
5. (MH "health status")
6. (MH "severity of illness index")
7. (MH "self-assessment")
8. TX outcome measure*
9. TX health outcome*
10. TX quality of life
11. TX health status
12. TX (end point* OR endpoint* OR end-point*)
13. TX (self-report* OR self report*)
14. TX functional outcome*
15. TI outcome*
16. OR/1-15
17. TX outcome*
18. TX measure*
19. TX assess*

20. (score* OR scoring)

21. TX index

22. TX indices

23. TX scale*

24. monitor*

25. OR/18-24

26. 17 AND 25

27. 16 OR 26

Block 2

28. TX interview*

29. TX experience*

30. TX qualitative

31. OR/28-30

Block 3

32. TX staff

33. TX professional*

34. TX personnel

35. OR/32-34

36. TX view*

37. TX opinion*

38. TX attitude*

39. OR/36-38

40. 35 N3 39

41. AND/27, 31, 40

CINAHL

Block 1

1. (MH "health status")
2. (MH "outcome assessment ")
3. (MH "quality of life")
4. (MH "health status")
5. (MH "severity of illness index")
6. (MH "self-assessment")
7. TX outcome measure*
8. TX health outcome*
9. TX quality of life
10. TX health status
11. TX (end point* OR endpoint* OR end-point*)
12. TX (self-report* OR self report*)
13. TX functional outcome*
14. TI outcome*
15. OR/1-14
16. TX outcome*
17. TX measure*
18. TX assess*

19. TX (score* OR scoring)

20. TX index

21. TX indices

22. TX scale*

23. TX monitor*

24. OR/17-23

25. 16 AND 24

26. 15 OR 25

Block 2

27. TX interview*

28. TX experience*

29. TX qualitative*

30. OR/27-29

Block 3

31. TX staff

32. TX professional*

33. TX personnel

34. OR/31-33

35. TX view*

36. TX opinion*

37. TX experience*

38. TX attitude*

39. OR/35-38

40. 34 N3 39

41. AND/26, 30, 40

PsychINFO

1. TX health status indicator

2. TX outcome assessment*

3. TX quality of life

4. TX health status

5. TX severity of illness index

6. TX self-assessment

7. TX outcome measure*

8. TX health outcome*

9. TX (end point* OR endpoint* OR end-point*)

10. TX (self-report* OR self report*)

11. TX functional outcome*

12. TI outcome*

13. OR/1-12

14. TX outcome*

15. TX measure*

16. TX assess*

17. TX (score* OR scoring)

18. TX index
19. TX indices
20. TX scale*
21. TX monitor*
22. OR/15-21
23. 14 AND 22
24. 13 OR 23

Block 2

25. TX interview*
26. TX experience*
27. TX qualitative*
28. OR/25-27

Block 3

29. TX staff
30. TX professional*
31. TX personnel
32. OR/29-31
33. TX view*
34. TX opinion*
35. TX experience*
36. TX attitude*
37. OR/33-36

38. 32 N3 37

39. AND/24, 28, 38

* Truncation; N3 proximity term; MH Mesh heading; TI Title; TX Text word

Appendix 3: Critical appraisal of studies included in the qualitative systematic review using CASP

Reference	Screening Q	Detailed Q							
	Aims and methods	Research design	Sampling	Data Collection	Reflexivity	Ethical issues	Data Analysis	Discussion of findings	Value
Bendtsen, 2003 (78)	Aims clearly stated and qualitative methods appropriate to examine attitudes towards value of HRQOL measurement.	Did not discuss how they decided which method to use.	All physicians that were present in the department were invited. No further discussion regarding why participants were selected and if they were the most appropriate.	Focus groups were used to collect the data in a library setting. Did not state why focus groups were chosen. An interview guide was used with 4 main themes. Focus groups were recorded and transcribed verbatim. Did not discuss saturation of data.	Potential bias in the formulation of questions or data collection was not discussed.	Lacking details on how the research was explained to participants, how researchers dealt with issues raised by the study (informed consent/confidentiality), or if ethical approval was sought.	Did not provide a description of the analysis process, how categories were derived from the data, how the data presented was selected. Sufficient data was presented to support findings and provided contradictory data but did not consider researcher bias on analysis.	Findings are explicit but researcher did not discuss credibility of findings.	Considered the value of the study and identified further research but did not address the generalizability of the findings.
Callaly, 2006 (105)	Aims clearly stated and qualitative methods	Qualitative methodology was used to learn more about	All clinicians within the Barwon Health mental health service were	Focus groups were undertaken unless participant preferred interview or	Potential bias in the formulation of questions or data	Lacking details on how the research was explained to participants	Analysis undertaken using grounded theory techniques. The interview guide	Findings are explicit. Discussed the credibility of findings as	Considered the value of the study, identified further

Reference	Screening Q	Detailed Q							
	Aims and methods	Research design	Sampling	Data Collection	Reflexivity	Ethical issues	Data Analysis	Discussion of findings	Value
	appropriate to explore attitudes towards the implementation and use of routine outcome measures.	clinician attitudes towards the utility and feasibility of using outcome measures.	invited to participate. This was one of four pilot agencies in Victoria. 83 out of 136 clinicians participated. Acknowledged that non-participants may have had different views.	couldn't attend the group. The setting of the interview was not specified. Did not state why focus groups were chosen. An interview guide was used with 8 questions. All discussions were recorded and 15 hours were transcribed. Did not discuss saturation of data.	collection was not discussed.	or how researchers dealt with issues raised by the study (informed consent/confidentiality). Ethics not sought as project was a quality assurance exercise.	provided framework for analysis. Did not provide a description of how the data presented was selected. Sufficient data was presented to support findings and provided contradictory data but did not consider researcher bias on analysis.	only one team member analysed the data. Identified that there was a possibility of bias towards clinicians with strong opinions and expressed a concern regarding differences between participants and non-participants views.	research and questioned the generalizability of the findings.
Cranley, 2004 (198)	Aims clearly stated and qualitative methods appropriate	The purpose of the article was to provide a theoretical perspective	The sample consisted of 29 nurses working in one of the wider study's participating	Informal semi-structured interviews were undertaken in a quiet corner due to busy working	Potential bias in the formulation of questions or data collection	Lacking details on how the research was explained to participants. Nurses signed	Content analysis was used. Categories were devised by assigning codes to the data. Data	Findings are explicit but researcher did not discuss credibility of	Considered the value of the study, identified further research and

Reference	Screening Q	Detailed Q							
	Aims and methods	Research design	Sampling	Data Collection	Reflexivity	Ethical issues	Data Analysis	Discussion of findings	Value
	to provide insight into how nurses integrate outcome assessment into practice.	into how nurse use assessment data to guide practice.	institutions. Interviews were held on different days to capture the team rotations. Does not state how many nurses were eligible for inclusion, how many refused to participate or if participants interviewed were the most appropriate.	environment. Did not state why interviews were chosen. Three open-ended questions were posed to each participant. Responses were hand recorded and field notes were taken after each interview. Collected for 8 weeks until saturation was reached.	was not discussed.	consent and confidentiality was assured. Ethical approval was sought as part of a wider study.	presented was based on two theoretical perspectives. Sufficient data was presented to support findings and provided contradictory data but did not consider researcher bias on analysis.	findings.	questioned the generalizability of the findings.
Dorwick, 2009 (199)	Aims clearly stated and qualitative methods appropriate to explore GPs	Did not discuss how they decided which method to use.	A sampling frame consisted of 38 general practices in 3 locations in England who were also taking part in	Open ended in-depth interviews were undertaken by three researchers primarily in GPs own surgeries. Did not state why	Considered bias in recruitment and data collection.	Lacking details on how the research was explained to participants, how researchers dealt with	Principles of constant comparison were used to analyse the data. Categories were derived using open, axial and	Findings are explicit. Discussed credibility of findings stating that there was a possibility of	Considered the value of the study, identified further research and questioned the

Reference	Screening Q	Detailed Q							
	Aims and methods	Research design	Sampling	Data Collection	Reflexivity	Ethical issues	Data Analysis	Discussion of findings	Value
	opinions of routine measurement.		quantitative study. Maximum variation approach employed. 34 GPs were interviewed. Stated that only interested GPs were likely to take part.	interviews were chosen. A topic guide was derived from the literature. Interviews were recorded and transcribed verbatim. Saturation of themes was not discussed.		issues raised by the study (informed consent/confidentiality). Ethical approval was stated.	selective coding. Tested thematic scheme and interpretation by reaching team consensus through iterative discussion. The data presented was selected by focusing on key themes that explained most of the data. Sufficient data was presented to support findings and considered deviant cases. Three researchers analysed the data to reduce bias in the interpretation of the data.	bias towards clinicians with strong opinions and that the researchers took care in avoiding the de-contextualisation of participant's words. Employed methods to increase credibility.	generalizability of the findings.
Dunckley,	Aims	Adopted	A sampling	Action research	Potential bias	Recruitment	Data from three	Findings are	Considered

Reference	Screening Q	Detailed Q							
	Aims and methods	Research design	Sampling	Data Collection	Reflexivity	Ethical issues	Data Analysis	Discussion of findings	Value
2005 (200)	clearly stated and qualitative methods appropriate to identify barriers and facilitators to implementing outcome measures.	action research as it identifies needs and defines problems while simultaneously devising methods of meeting those needs and readdressing problems around service provision.	frame consisted of 28 nursing home staff and 23 clinical hospice staff. 8 nurses, 1 doctor and 6 healthcare assistants participated in the pre-implementation interview. Participants from 3 nursing homes and one hospice declined to take part in the second interview and did not provide a reason for their decision. Staff completed a diary and attended action research	including collaborative meetings, interviews and diaries were undertaken with staff in two contrasting settings. Did not state the setting of the interviews. The discussion focused on experiences of using the POS. Did not explicitly state if interviews were recorded or transcribed. Saturation of themes was not discussed.	in the formulation of questions or data collection was not discussed.	packs were provided to all eligible staff. Participants signed consent and were assigned random IDs to maintain anonymity. Ethics approval stated.	sources were analysed. Text was coded and grouped thematically by content. The data presented was selected for illustrative purposes. Sufficient data was presented to support findings and provided contradictory data but did not consider researcher bias on analysis.	explicit but researcher did not discuss credibility of findings.	the value of the study, identified further research and questioned the generalizability of the findings.

Reference	Screening Q	Detailed Q							
	Aims and methods	Research design	Sampling	Data Collection	Reflexivity	Ethical issues	Data Analysis	Discussion of findings	Value
			collaborative meetings. It is unclear whether these were the same participants that undertook the interview.						
Eischens, 1998 (201)	Aims clearly stated and qualitative methods appropriate to assess if QOL evaluations are useful in designing care plans.	Did not discuss how they decided which method to use.	A sampling frame consisted of 9 home care nurses. 8 participated in an interview. One nurse refused to take part but did not provide a reason for this decision.	Interviews were undertaken with nurses. The setting of the interview was not specified. Did not state why interviews were chosen. An interview guide was developed with 12 questions. Field notes taken and later transcribed. Did not discuss	Potential bias in the formulation of questions or data collection was not discussed.	Researchers presented the project to all nurses. Lacking details on how researchers dealt with issues raised by the study (informed consent/confidentiality). Ethics approval stated.	Did not provide a description of the analysis process, how categories were derived from the data, or how the data presented was selected. Compared two measures and considered researcher bias involving 3 analysts.	Findings are explicit. Discussed the credibility of finding as the researcher highlighted the preliminary nature of the findings given the size of the sample.	Considered the value of the study, identified further research and questioned the generalizability of the findings.

Reference	Screening Q	Detailed Q								
		Aims and methods	Research design	Sampling	Data Collection	Reflexivity	Ethical issues	Data Analysis	Discussion of findings	Value
					saturation of data.					
Hughes, 2003 (202)	Aims clearly stated and qualitative methods appropriate to elicit professional's experiences of using outcome measures.	Qualitative methodology was used to learn more about professional's views and experiences of using outcome measures.	A sampling frame consisted of 26 professionals. 22 participated in an interview. Four declined due to sick leave, analysis of data had not been completed or due to time constraints.	Telephone interviews undertaken with professionals. Did not state why interviews were chosen. An interview guide was developed based on previous research. Verbatim notes were taken thought out. Did not discuss saturation of data.	Potential bias in the formulation of questions or data collection was not discussed.	Researchers contacted eligible participants by email and phone. Lacking details on consent procedures but confidentiality was assured. Ethical approval was sought as part of a wider study.	Notes were coded and thematically analysed by content and categorised according to common thematic grouping. Quotations refer to telephone interview notes but did not link quotes to participants. Sufficient data was presented to support findings and provided contradictory data but did not consider researcher bias on analysis.	Findings are explicit. The researcher highlighted the preliminary nature of the findings given the purposive sample however they did not explicitly discuss credibility of findings.	Considered the value of the study, identified further research and questioned the generalizability of the findings.	

Reference	Screening Q	Detailed Q							
	Aims and methods	Research design	Sampling	Data Collection	Reflexivity	Ethical issues	Data Analysis	Discussion of findings	Value
Hughes, 2004 (203)	Aims clearly stated and qualitative methods appropriate to understand the implementation of an outcome measure.	Qualitative methodology was used to learn more about professional's views and experiences of using outcome measures.	A sampling frame consisted of 25 palliative care settings. 15 participated in the study in two geographical regions. 13 staff interviewed to understand the implementation of the POS. No further discussion regarding why participants were selected and if they were the most appropriate.	Semi-structured interviews were undertaken with professionals in each organisational setting. Did not state why interviews were chosen. Does not provide details on interview guide. Interviews were recorded and transcribed verbatim. Saturation of themes was not discussed.	Potential bias in the formulation of questions or data collection was not discussed.	Lacking details on how the research was explained to participants or how researchers dealt with issues raised by the study (informed consent/confidentiality). Ethics approval stated.	Data was coded and sorted. Stated that analysis was completed according to established procedures but lacking details on how categories were derived from the data. Data presented was selected for illustrative purposes but was not an exhaustive representation of the data set. Sufficient data was presented to support findings and provided contradictory data but did not consider researcher bias on	Findings are explicit but researcher did not discuss credibility of findings.	Considered the value of the study and identified further research but did not address the generalizability of the findings.

Reference	Screening Q	Detailed Q							
	Aims and methods	Research design	Sampling	Data Collection	Reflexivity	Ethical issues	Data Analysis	Discussion of findings	Value
							analysis.		
Kettis-Lindblad, 2007 (204)	Aims clearly stated and qualitative methods appropriate to explore oncologist's views of QOL assessments to support the consultation.	Qualitative methodology was used to learn more about complex healthcare interventions and their potential effects.	A sampling frame consisted of 8 oncologists in two hospital settings. 6 participated in an interview. No further discussion regarding why participants were selected and if they were the most appropriate.	Interviews were undertaken with oncologists. The setting of the interview was not specified. Did not state why interviews were chosen. An interview guide was used, and interviews were recorded and transcribed verbatim. Saturation of themes referred to in relation to patient interviews but not professional interviews.	Potential bias in the formulation of questions or data collection was not discussed.	Lacking details on how the research was explained to participants, how researchers dealt with issues raised by the study (informed consent/confidentiality). Ethics approval stated.	Interpretative approach adopted. Data was coded and emerging themes were refined iteratively into categories and thematically analysed by content. Data presented was selected for illustrative purposes. Sufficient data was presented to support findings but did not provide contradictory data. Considered researcher bias involving 2	Findings are explicit but researcher did not explicitly discuss credibility of findings. However methods were employed to increase credibility.	Considered the value of the study, identified further research and questioned the generalizability of the findings.

Reference	Screening Q	Detailed Q							
	Aims and methods	Research design	Sampling	Data Collection	Reflexivity	Ethical issues	Data Analysis	Discussion of findings	Value
							analysts.		
Mason, 2008 (205)	Aims clearly stated and qualitative methods appropriate to access the perceptions of healthcare professionals towards screening using the EPDS.	Qualitative methodology was used to understand beliefs behind attitudes which a survey would fail to do.	A sampling frame consisted of 22 health professionals who routinely screen women for PND. 19 participated in an interview. No further discussion regarding why participants were selected and if they were the most appropriate.	Semi-structured interviews were undertaken in a private room in an NHS clinic. Did not state why interviews were chosen. A topic guide was used, and interviews were recorded and transcribed verbatim. Saturation of themes was not discussed.	Potential bias in the formulation of questions or data collection was not discussed.	Written information was provided to all eligible staff. Participants signed consent but lacking details on confidentiality. Ethics approval stated.	Interpretative phenomenological approach was used for the analysis. Scripts were double checked for accuracy. Read and reread to identify emerging themes and categories. Iterative process so as a new theme emerged all scripts were rechecked. Did not provide a description of how the data presented was selected. Sufficient data was presented to	Findings are explicit but researcher did not explicitly discuss credibility of findings. However methods were employed to increase credibility.	Considered the value of the study, identified further research and questioned the generalizability of the findings.

Reference	Screening Q	Detailed Q							
	Aims and methods	Research design	Sampling	Data Collection	Reflexivity	Ethical issues	Data Analysis	Discussion of findings	Value
							support findings and provided contradictory data. Considered researcher bias involving 2 analysts.		
Meehan, 2006 (106)	Aims clearly stated and qualitative methods appropriate to explore clinician's reactions to the introduction and utility of outcome measures.	Qualitative methodology was used to learn more about clinician's reactions to outcome data.	Aimed to get a representative sample across broad range of services. 34 focus groups were held. No further discussion regarding why participants were selected and if they were the most appropriate.	Focus groups were undertaken using semi-structured format. The setting of the groups was not specified. Did not state why focus groups were chosen. An interview guide was used and extensive notes taken for all groups. 15 interviews were recorded and transcribed	Potential bias in the formulation of questions or data collection was not discussed.	Information was provided to all service leaders on aims and format of the discussion. Lacking details regarding how researchers dealt with issues raised by the study (informed consent/confidentiality). Ethics	Content analysis was used for the analysis. Data was reviewed, coded and categorised. Did not provide a description of how the data presented was selected. Sufficient data was presented to support findings and provided contradictory data. Considered researcher bias	Findings are explicit. Discussed credibility of findings as stated that there was a possibility of bias towards clinicians with strong opinions and localised circumstances. Also stated that research team's interpretation	Considered the value of the study, identified further research and questioned the generalizability of the findings.

Reference	Screening Q	Detailed Q							
	Aims and methods	Research design	Sampling	Data Collection	Reflexivity	Ethical issues	Data Analysis	Discussion of findings	Value
				verbatim as researchers were conscious that rich information was provided once the tape was turned off. Saturation of themes was not discussed but included a large sample.		approval stated.	involving numerous analysts.	could be bias, however they tried to eliminate this by involving numerous members in the analysis process.	
Mitchell, 2011 (206)	Aims clearly stated and qualitative methods appropriate to explore the clinical utility of NICE guideline and QOF for	Did not discuss how they decided which method to use.	A sampling frame consisted of 26 practices in south Yorkshire. Five responded and maximum variation approach was used to choose four diverse practices.	Focus groups were undertaken (min 8 and max 10) comprising of 38 participants. The setting of the groups was not specified. Justified the use of focus groups. A topic guide was developed from a literature review and consultation	Potential research bias was considered in the collection and analysis of the data.	Lacking details on how researchers dealt with issues regarding confidentiality. Informed consent was obtained and ethical approval was stated.	The analysis was iterative, thematic and self-conscious. Data was coded, grouped into themes and compared across groups. Did not provide a description of how the data presented was selected.	Findings are explicit. Discussed credibility of findings stating that focus groups may have led to misrepresentation of views. The influence of a GP researcher may	Considered the value of the study, identified further research and questioned the generalizability of the findings.

Reference	Screening Q	Detailed Q							
	Aims and methods	Research design	Sampling	Data Collection	Reflexivity	Ethical issues	Data Analysis	Discussion of findings	Value
	depression.			with experts, and was piloting. A trained academic GP facilitated interviews along with an additional person to take observational notes. Groups were recorded and transcribed verbatim. Saturation of themes was not discussed.			Sufficient data was presented to support findings, provided contradictory data, considered researcher bias involving numerous analysts and ensuring analysis meetings challenged approach.	have influenced group dynamic, however authors tried to reduce bias by ensuring a self-conscious process and multidisciplinary analysis.	
Slater, 2005 (207)	Aims clearly stated and qualitative methods appropriate to elicit experience of using the POS and its	Qualitative methodology was used to produce rich and insightful information.	A sampling frame consisted of 9 staff in one day hospice. 8 agreed to participate. One staff member was not available on that	Focus group undertaken. The setting of the groups was not specified. Did not state why focus groups were chosen. A topic guide was developed through	Potential bias in the formulation of questions was not discussed. Acknowledged that independent researcher	Information was provided to all staff. Consent issues were discussed prior to signing the form, confidentiality was assured by	Interpretative phenomenological approach was used. Data was reviewed, coded and categorised through an iterative process. Did not provide a description of	Findings are explicit and researchers discussed the credibility of findings which was limited by the small sample. However	Considered the value of the study, identified further research and questioned the generalizability of the

Reference	Screening Q	Detailed Q							
	Aims and methods	Research design	Sampling	Data Collection	Reflexivity	Ethical issues	Data Analysis	Discussion of findings	Value
	usefulness.		day.	discussion with researcher and principal investigator. Groups were recorded and transcribed. Saturation of themes was not discussed.	facilitated focus groups to minimise bias in data collection.	anonymising the data and ethical approval was stated.	how the data presented was selected. Sufficient data was presented to support findings and provided contradictory data. Considered researcher bias involving 2 analysts.	methods were employed to increase credibility.	findings.
Tavabie, 2009 (208)	Aims clearly stated and qualitative methods appropriate to investigate the impact of mental health questionnaires on	Qualitative methodology was used to investigate the impact of questionnaires on clinician.	A sampling frame consisted of 21 GPs in 4 practices. Four GPs were used for a pilot study and 16 additional GPs participated. One GP refused to part take and did not provide a reason for this	Focus groups and interviews were undertaken. The setting of the groups was not specified. Both methods were used to triangulate data. Topic guide was developed and allowed to evolve throughout the process.	Recognised that the researcher was a GP in one of the practices which may have introduced bias in data collection.	Lacking details on how the research was explained to participants and whether informed consent was sought. Confidentiality was assured and ethics approval	Inductive principles of grounded theory were applied using a constant comparative analysis and iterative process. Initial codes were grouped into categories from which themes emerged. Did not	Findings are explicit. Discussed credibility of findings by using respondent validation, triangulation and multiple analysts.	Considered the value of the study, identified further research and questioned the generalizability of the findings.

Reference	Screening Q	Detailed Q							
	Aims and methods	Research design	Sampling	Data Collection	Reflexivity	Ethical issues	Data Analysis	Discussion of findings	Value
	views of GPs when dealing with depression.		decision.	Discussions were recorded and transcribed verbatim. Data collection continued until saturation was reached. Respondent checking ensured reliability.		stated.	provide a description of how the data presented was selected. Sufficient data was presented to support findings and provided contradictory data. Considered researcher bias involving 2 analysts.		
Unsworth, 2011 (209)	Aims clearly stated and qualitative methods appropriate to elicit professional's perceptions on the use	Qualitative methodology was used to adopt and naturalistic method of enquiry.	Unclear details on the sampling frame. Four therapists who were experienced in using CORE-NET and five therapists who had just begun using CORE-	Two focus groups were undertaken with therapists by two different researchers. The setting of the groups was not specified. Did not state why focus groups were chosen. A semi-	Potential research bias in the data collection as roles was considered. A reflective research diary was completed	Information was provided to all professionals by letter. Lacking details on how researchers dealt with issues regarding	Inductive approach applied using in vivo coding which was continuously revised and refined throughout the process. The data presented was selected to	Findings are explicit and researchers discussed the credibility of finding which was limited by the small sample. However methods were	Considered the value of the study, identified further research and questioned the generalizability of the

Reference	Screening Q	Detailed Q							
	Aims and methods	Research design	Sampling	Data Collection	Reflexivity	Ethical issues	Data Analysis	Discussion of findings	Value
	of CORE-NET in practice.		NET participated. No further discussion regarding why participants were selected and if they were the most appropriate.	structured interview guide was used. Groups were recorded and transcribed. Saturation of themes was not discussed. Member checking ensured reliability.	throughout.	confidentiality. Informed consent was obtained and ethical approval was stated.	convey the core theme or essence of a category. Sufficient data was presented to support findings and provided contradictory data. Considered researcher bias as a supervisor checked for accuracy in the analysis.	employed to increase credibility.	findings.
Wressle, 2003 (210)	Aims clearly stated and qualitative methods appropriate to investigate the usefulness of the	Did not discuss how they decided which method to use.	A sampling frame consisted of 7 professionals in a multidisciplinary team and all took part in the study. Discussed changes to the team during	Interviews undertaken with team members. The setting of the groups was not specified. Did not state why interviews were chosen. A discussion guide was used.	Potential research bias was considered in the collection and analysis of the data.	Lacking details on how the research was explained to participants and how researchers dealt with issues regarding confidentiality.	Principles of grounded theory were applied going from open coding to axial coding. Responses examined for patterns or trends. Data presented was selected for	Findings are explicit but researcher did not discuss credibility of findings.	Considered the value of the study, identified further research and questioned the generalizability of the

Reference	Screening Q	Detailed Q							
	Aims and methods	Research design	Sampling	Data Collection	Reflexivity	Ethical issues	Data Analysis	Discussion of findings	Value
	COPM		study period. No further discussion regarding why participants were selected and if they were the most appropriate.	Interviews were recorded and transcribed. Saturation of themes was not discussed.		Informed consent was obtained and ethical approval was stated.	illustrative purposes. Sufficient data was presented to support findings, provided contradictory data but did not consider researcher bias on analysis		findings.

Appendix 4: Themes identified as barriers and facilitators to the use of PROMs within each study

Reference	Study design	Theme 1- Practical considerations		Theme 2- Valuing the data		Theme 3- Methodological considerations		Theme 4- Impact on patient care		Study characteristics
		Barrier	Facilitator	Barrier	Facilitator	Barrier	Facilitator	Barrier	Facilitator	
Bendtsen, 2003 (78)	Focus: Use of information Level: Individual Collection: Computer	-	Technology	-	-	-	Interpretation	No clinical value	Promotes quality improvement, Positive indirect effects	Setting: Secondary Professional: Clinician Function: Care management Healthcare issue: COPD
Callaly, 2006 (105)	Focus: Implementation and use Level: Individual and group Collection: Computer	Workload, Administration, Clear guidelines, Training, Technology	Training, Support, Technology, Involvement of management / use of data	Open to feedback and change, Clarity of objectives	-	Interpretation, Validity	Interpretation	No clinical value, Negative indirect effects	Promotes quality improvement	Setting: Mixed Professional: Mixed Function: Performance measure & care management Healthcare issue: Mental Health
Cranley, 2004 (198)	Focus: Use of information Level: Individual	-	-	Open to feedback and change	-	-	-	No clinical value	Promotes quality improvement	Setting: Secondary Professional: Nurse Function: Care

	Collection: Paper									management Healthcare issue: Acute care
Dorwick, 2009 (199)	Focus: Use of information Level: Individual Collection: Paper	-	-	Open to feedback and change, Clarity of objectives	-	Validity	-	Negative indirect effects	Promotes quality improvement, Positive indirect effects	Setting: Primary Professional: Clinician Function: Care management Healthcare issue: Mental Health
Dunckley, 2005 (200)	Focus: Implementation and use Level: Unclear Collection: Paper	Workload, Collaboration, Clear guidelines, Involvement of management/ use of data, Training/ familiarisation, Support	Administration, Involvement of management / use of data	-	-	Sensitivity	-	No clinical value, Negative indirect effects	Promotes quality improvement, Positive indirect effects	Setting: Mixed Professional: Mixed Function: Unclear Healthcare issue: Palliative care
Eischens, 1998 (201)	Focus: Implementation and use Level:	Administration, Clear guidelines	Administration	-	-	-	-	-	Promotes quality improvement	Setting: Hospice Professional: Nurse

	Individual Collection: Paper									Function: Care management Healthcare issue: Palliative care
Hughes, 2003 (202)	Focus: Implementation and use Level: Individual Collection: Paper	Workload, Administration	Administration, Flexibility in administration, Training/familiarisation	-	-	Validity, Sensitivity	-	No clinical value, Negative indirect effects	Promotes quality improvement	Setting: Unclear Professional: Unclear Function: Care management Healthcare issue: Palliative care
Hughes, 2004 (203)	Focus: Implementation and use Level: Individual Collection: Paper	Workload, Administration, Collaboration, Training	-	Clarity of objectives	-	Validity	-	Negative indirect effects	Promotes quality improvement, Positive indirect effects	Setting: Mixed Professional: Mixed Function: Care management Healthcare issue: Palliative care

Kettis-Lindblad, 2007 (204)	Focus: Implementation and use Level: Individual Collection: Computer	Workload, Clear guidelines, Involvement of management/ use of data, Training	Streamlines workload, Technology	Open to feedback and change	-	-	Interpretation		Promotes quality improvement	Setting: Secondary Professional: Clinician Function: Care management Healthcare issue: Oncology
Mason, 2008 (205)	Focus: Implementation and use Level: Individual Collection: Paper	Workload, Administration, Clear guidelines	Training, Support	-	-	Validity	-	No clinical value, Negative indirect effects	Promotes quality improvement, Positive indirect effects	Setting: Primary Professional: Mixed Function: Screening Healthcare issue: Mental Health
Meehan, 2006 (106)	Focus: Implementation and use Level: Individual and group Collection: Computer	Workload, Administration, Collaboration, Training, Involvement of management/ use of data, Technology, Support	Training/ familiarisation, Support, Involvement of management / use of data	Open to feedback and change, Clarity of objectives	-	Validity	-	No clinical value, Negative indirect effects	Promotes quality improvement	Setting: Mixed Professional: Mixed Function: Performance measure & care management Healthcare issue: Mental Health

Mitchell, 2011 (206)	Focus: Implementation and use Level: Individual Collection: Paper	Workload, Administration, Clear guidelines, Training, Support	-	Open to feedback and change	-	Interpretation Validity	-	No clinical value, Negative indirect effects	-	Setting: Primary Professional: Mixed Function: Care management Healthcare issue: Mental Health
Slater, 2005 (207)	Focus: Implementation and use Level: Individual Collection: Paper	Workload, Administration Involvement of management/ use of data, Support	Support	Open to feedback and change, Clarity of objectives	-	Interpretation Validity Sensitivity	-	No clinical value, Negative indirect effects	Promotes quality improvement	Setting: Hospice Professional: Mixed Function: Screening and care management Healthcare issue: Palliative care
Tavabie, 2009 (208)	Focus: Use of information Level: Individual Collection: Computer	Administration, Collaboration, Support	Streamlines workload	Open to feedback and change	-	Validity	-	Negative indirect effects	Promotes quality improvement Positive indirect effects	Setting: Primary Professional: Clinician Function: Screening and care management Healthcare issue: Mental

										Health
Unsworth, 2011 (209)	Focus: Implementation and use Level: Individual Collection: Computer	Training, Technology	Streamlines workload, Flexibility in administration, Training	Open to feedback and change, Clarity of objectives	-	-	-	-	Promotes quality improvement Positive indirect effects	Setting: Primary Professional: Therapist Function: Care management Healthcare issue: Mental Health
Wressle, 2003 (210)	Focus: Use of information Level: Individual Collection: Paper	Clear guidelines	-	Open to feedback and change	-	Validity	-	-	Promotes quality improvement	Setting: Secondary Professional: Mixed Function: Care management Healthcare issue: RA

Appendix 5: Feedback report provided to surgeons in the trial

Appendix 6: Educational session provided to surgeons in the trial

Appendix 7: Training session provided to data collectors

Appendix 8: Information sheet provided to patients in the trial

Appendix 9: Study handbook provided to the data collectors

Appendix 10: Pre-operative questionnaire

Appendix 11: Post-operative questionnaire

Appendix 12: Pre-feedback differences in age, gender and scores between patients included in the study, excluded from the analysis and excluded from the study.

Characteristic	Patients in the analysis (n=624)	Patients excluded from the analysis due to non-response (n=108)	Patients excluded from the study due to surgeons not meeting inclusion criteria (n=269)
Age, mean (SD)	66.3 (11.28)	63.7 (11.6)	66.1 (12.1)
Male, n (%)	333 (53%)	49 (46%)	131 (50%)
Health status, n (%)			
Excellent	75 (12%)	9 (9%)	23 (9%)
V. Good	218 (36%)	31 (30%)	81 (31%)
Good	232 (38%)	49 (47%)	107 (41%)
Fair	76 (13%)	12 (11%)	46 (18%)
Poor	7 (1%)	4 (4%)	4 (2%)
Duration of symptoms, n (%)			
<1 year	117 (19%)	18 (17%)	36 (14%)
1-5 years	411 (66%)	68 (63%)	173 (65%)
6-10 years	55 (9%)	13 (12%)	31 (12%)
>10 years	41 (7%)	8 (7%)	24 (9%)

Characteristic	Patients in the analysis (n=624)	Patients excluded from the analysis due to non-response (n=108)	Patients excluded from the study due to surgeons not meeting inclusion criteria (n=269)
OHS pre-op, mean (SD)	19.6 (9.0)	17.4 (8.6)	19.3 (9.0)
OHS post-op, mean (SD)	40.8 (7.4)	NA	38.6 (8.5)
HOOS pre-op, mean (SD)	17.3 (8.5)	15.7 (8.3)	16.8 (8.6)
HOOS post-op, mean (SD)	36.0 (7.8)	NA	34.6 (8.0)
EQ5D pre-op, mean (SD)	0.4 (0.3)	0.3 (0.3)	0.4 (0.3)
EQ5D post-op, mean (SD)	0.85 (0.2)	NA	0.78 (0.01)

Appendix 13: Pre-feedback patient characteristics between arms.

Characteristic	Total (n=732)	Control group (n=332)	Intervention group (n=400)
Age, mean (SD)	65.9 (11.4)	66.3 (10.8)	65.5 (11.8)
Men, n (%)	382 (52)	169 (51)	213 (53)
Health status, n (%)			
Excellent	84 (12)	27 (8)	57 (15)
V. Good	249 (35)	120 (37)	129 (33)
Good	281 (39)	136 (42)	145 (37)
Fair	88 (12)	34 (11)	54 (14)
Poor	11 (2)	5 (2)	6 (2)
Duration of symptoms, n (%)			
<1 year	135 (18)	58 (18%)	77 (19%)
1-5 years	479 (66)	229 (69%)	250 (63%)
6-10 years	68 (9)	26 (8%)	42 (10%)
>10 years	49 (7)	19 (6%)	30 (7%)
OHS pre-op, mean (SD)	19.3 (9)	19.5 (8.8)	19.1 (9.2)
OHS post-op, mean (SD)	40.8 (7.4)	40.8 (7.2)	40.8 (7.5)
HOOS pre-op, mean (SD)	17.1 (8.5)	17.4 (8.4)	16.8 (8.6)
HOOS post-op, mean (SD)	35.9 (7.8)	35.8 (7.8)	36.1 (7.8)

Characteristic	Total (n=732)	Control group (n=332)	Intervention group (n=400)
EQ5D pre-op, mean (SD)	0.39 (0.3)	0.42 (0.3)	0.38 (0.3)
EQ5D post-op, mean (SD)	0.85 (0.2)	0.84 (0.2)	0.85 (0.2)

Appendix 14: Discussion guide for interviews

Firstly, I would like to thank you for taking the time to talk to me. My name is Maria Boyce, I am a researcher based in UCC. I would like to explain the background behind this research and rationale for the interview. I will then ask you to sign the consent form before we commence the interview.

I am going to use the term PROMs throughout the interview. This stands for Patient-reported outcome measures which are questionnaires that assess patients' health including: symptoms, function, well-being, health-related quality of life (HRQOL) and other health-related constructs.

As you may know, the NHS introduced a national programme in 2009 which made the collection of PROMs a mandatory requirement for audit. Therefore, every patient that receives a hip or knee replacement surgery is asked to complete a questionnaire to assess their pain and function before and six months after their operation. The results are compared at a hospital level and are publically reported online to inform patient choice. They intend to extend the use of PROMs to other areas such as mental health, oncology and some chronic conditions, and they have also plans to link payments to results. This programme is stimulating much interest from policy makers internationally. However, there is little empirical evidence on the use of PROMs as a performance measure and so this study is the first to evaluate the usefulness of this strategy. Furthermore, professionals are the target of such an initiative but there has being no attempt to evaluate their views on the usefulness of such data and therefore this is the focus of this interview.

I am undertaking an interview with every surgeon in the feedback arm of the PROFILE trial. You received benchmarked feedback after Christmas which was based on PROMs data. I would like to establish your views on the collection and

value of such data. The interview should last about 30-40 minutes. I would just like to check a few things before we get started.

- Would you mind if I record this interview? Anything we discuss will be confidential and your identity will remain anonymous on any reports or publications. Finally you can stop the interview at any point, if you wish. Do you have any questions for me before we get started?
- Sign consent and give copy

Background

- Firstly, could you tell me about your experience with the use of PROMs?
- Have you (or the hospital you work in) collected PROMs before we begun this trial?
 - YES
 - What measures do you collect?
 - How do you use this data (dissemination: reports, meetings)?
 - How do you think this information should be used?
 - No
 - Can you explain to me any QI initiative in which you involved the patient?
 - What is your experience of QI initiatives in the hospitals you work in?

Attitudes

- What are your views on the collection and use of PROMs?
 - In particular, what are your opinions on the use of PROMs as a QI tool?

- How would you feel if this data was used :
 - As a clinical governance tool in the hospital(s) you work in?
 - To inform patient choice by publically reporting the data?
 - To inform purchasers decisions?
 - To link payment to results?
- How do you think PROMs should be used?
- Would you like to receive regular feedback reports?

Methodological issues

- Moving on, one of the things I am particularly interested in is the thought process when you read the report? Could you explain this to me?
- In particular, what factors do you think affected the results (patient, surgeon, hospital)?
 - Do you agree with the findings?
 - What are your views on patients reporting on these issues?
- Did you understand the feedback report?
 - What information did you find useful?
 - Was there anything you did not understand?
 - How could we improve the report?
 - Did you find the feedback clinically meaningful?
- Would you use these measures to detect a change in outcomes over time or across surgeons?

Impact

- In theory, we assume that providing surgeons with benchmarked feedback will promote changes in patient care. How do you think this happens in practice?
 - Could you describe any changes you made/would like to make based on these findings?
 - Do you think this feedback would stimulate further research/audit?
 - Steps taken to implement changes?
- If no, what factors may prevent change?
- How else could PROMs feedback impact on care or practice?
- There is a debate about the level at which PROMs should be fed back (surgeon/hospital), could you describe your opinion on this matter?

Practical issues

- From a practical point of view, is the routine collection of PROMs is feasible?
 - Administration, coding, analysis, interpretation?
- What would facilitate the collection and use of PROMs?
 - Support required to collect and appropriately use the information?
 - Guidelines or training/educational needs?
 - Role of technology?

Thanks for sharing your views and experience with me. Have you any additional questions or anything else to add before we finish?

Appendix 15: Examples of monthly newsletters provided to data collectors

Appendix 16: Published papers