RECOMMENDATIONS TO OECD MINISTERS OF HEALTH FROM THE HIGH LEVEL REFLECTION GROUP ON THE FUTURE OF HEALTH STATISTICS

Strengthening the international comparison of health system performance through patient-reported indicators

January 2017

OECD
BETTER POLICIES FOR BETTER LIVES

RECOMMENDATIONS TO OECD MINISTERS OF HEALTH FROM THE HIGH LEVEL REFLECTION GROUP ON THE FUTURE OF HEALTH STATISTICS

Strengthening the international comparison of health system performance through patient-reported indicators

Members of the High Level Reflection Group

The group convened at the OECD, Paris on 4th May 2015
and on 21st September 2015.

Donald Berwick	Institute for Healthcare Improvement, United States
Nick Black	London School of Hygiene & Tropical Medicine, United Kingdom
David Cullen	Department of Health, Australia
Jens Deerberg-Wittram	International Consortium for Health Outcomes Measurement
Laurent Degos	University of Paris Diderot-Hôpital Saint Louis, Pasteur Institute, France
Brent Diverty	Canadian Institute for Health Information
Arnold Epstein	U.S. Department of Health and Human Services
Marie-Paule Kieny	World Health Organization
Bjørn-Inge Larsen	Ministry of Health and Care Services, Norway (*Chair of the OECD Health Committee*)
Eric Schneider	The Commonwealth Fund, United States
Stefan Schreck	Directorate General for Health and Food Safety, Luxembourg
Steven Sheingold	U.S. Department of Health and Human Services
Olivia Wigzell	Swedish Council for Health Technology Assessment and Assessment of Social Services

Foreword

The OECD has historically played a leading role in measuring health system performance. Data generated by health systems, however, are too concentrated on health system inputs and activities. There remain substantive gaps in what is known about the experience of patients, and the outcomes of care from the patient's point of view. Opportunities for cross-country comparative analysis of outcomes are also very limited, hampering the capacity of policy makers to gain new knowledge that would help them provide health services shaped around patients' needs. Such information is pivotal to delivering health services that are truly responsive to patients. This is a major gap in international health statistics requiring urgent attention.

To address these critical information deficits and provide directions to the OECD Health Committee on future work in this area, the OECD convened a High Level Reflection Group (HLRG) on Health Statistics, composed of leading figures in measuring and driving health performance improvement across OECD countries, and chaired by the Chair of the Health Committee. The HLRG was asked provide advice on how the OECD could collect and report internationally comparative data that would present a more comprehensive picture of health system performance.

At its first meeting on 4 May 2015, the HLRG discussed options to improve the collection and reporting of health care outcomes across OECD populations. The HLRG held its second meeting on 21 September 2015. The meeting discussion provided advice to the OECD Secretariat on the implementation of standardised, validated instruments for the collection and reporting of patient-reported indicators of health system performance. This report presents the Recommendations of the HLRG concerning the future of health statistics, for OECD Ministers of Health to consider.

ACKNOWLEDGEMENTS

The High Level Reflection Group (HLRG) was convened by Francesca Colombo, Mark Pearson and Stefano Scarpetta from the OECD Directorate of Employment, Labour and Social Affairs.

The writing and production of this report was co-ordinated by Ian Forde. The other members of the Secretariat supporting the HLRG were Carol Nader, Niek Klazinga and Luke Slawomirski from the OECD's Health Division, and Philip van der Wees from Radboud University Nijmegen.

Thanks also go to Marlène Mohier and Lucy Hullet and for their editorial input and oversight during publication and to Isabelle Vallard, Duniya Dedeyn and Luckasz Lech for logistical assistance.

Table of contents

Figures

Tables

ACRONYMS AND ABBREVIATIONS

AMI	Acute myocardial infarction
CAHPS	Consumer Assessment of Healthcare Providers and Systems
CIHI	Canadian Institute for Health Information
COPD	Chronic obstructive pulmonary disease
CTM	Care Transitions Measure
DRG	Diagnosis Related Group
GP	General Practitioner
HCQI	Health Care Quality Indicator
HHFT	Hampshire Hospitals Foundation Trust (England)
ICHOM	International Consortium for Health Outcome Measurement
NHS	National Health Service
NQF	National Quality Forum (United States)
NQR	National Quality Registry
NRS	Numeric Rating Scale
P4P	Pay for performance
PREM	Patient-reported experience measure
PRO(M)	Patient-reported outcome (measure)
PROMIS	Patient-reported Outcomes Measurement Information System
QLQ	Quality of Life Questionnaire
VAS	Visual Analogue Scale

Key recommendations

Globally, health systems need better information on the value and outcomes they produce. Critical knowledge gaps continue to hamper efforts to better understand and improve health system performance, particularly for the increasing share of the population that live with complex, chronic conditions.

Historically, health systems have not done enough to assess health care quality and outcomes from the perspective of those most concerned – patients themselves and their carers. Addressing this lack of patient-reported indicators of performance is an urgent need.

Whilst it is promising that several patient-reported indicators have been developed, each health system is currently pursuing its own path. If each country continues to do its own thing on patient-reported performance, opportunities to identify excellence, support poor performers and drive improvements across the board will be missed. There is clear and substantial benefit, therefore, from standardising these performance indicators across countries.

The OECD has historically played a leading role in measuring health system performance, and is well positioned to develop, collect and analyse patient-reported indicators for cross-country comparison. The OECD is also ideally placed to convene the political will to make this much-needed change happen, and provide the forum to ensure that patient-reported performance benchmarks are applied to drive improvement.

1. General principles

- OECD work to extend and deepen the benchmarking of health system performance should focus on collecting patient-reported indicators at a disease level, sector level, health-service level, and whole-system level.[1] In each case, the focus should be on enabling international comparison.

- The prioritisation of patient-reported indicators for development should be guided by the following criteria:

  - Work should initially focus on clinical areas where the OECD already collects other data, such as prevalence of risk factors, indicators of need, activity volumes or survival estimates. Such complementary data should be used to place patient-reported performance in the broader health system context.

  - Work should begin with validated indicators that countries are already using and, where possible, seek to accelerate international adoption and/or harmonisation across countries.

  - Prioritisation should also take patients' priorities into account, identified through surveys, focus groups or other means.

- Where valid patient-reported indicators do not yet exist for priority diseases, sectors or services, new indicators and patient surveys should be developed.

- All indicators should be formally assessed and piloted in different languages and settings, to ensure feasibility, utility and validity for the purposes of international comparison.

- The OECD should use its established structures, principally the Health Care Quality Indicators Expert Group and the Health Committee, to guide the prioritisation of patient-reported indicators for development, and the technical assessment of feasibility, utility and validity for the purposes of international comparison.

- The OECD should explore collaboration with other international organisations, such as the World Health Organisation, the European Commission, the Commonwealth Fund and the International Consortium for Health Outcomes Measurement (ICHOM), work to extend and deepen the benchmarking of patient-reported performance indicators.

2. International benchmarking of patient-reported experience measures (PREMs)

- Initially, the OECD should build on its collaboration with the Commonwealth Fund to benchmark PREMs in ambulatory care.

Currently, benchmarks are available for 19 countries. This number should be expanded.

- Subsequently, the OECD should work with countries to extend PREMs to clinical areas that have received little attention to date: mental health care, long-term care, palliative care, emergency care, informal care and preventive care.

- The OECD should develop surveys of patient experience that assess neglected aspects of care: co-ordination for individuals with chronic conditions, and patient safety.

3. International benchmarking of patient-reported outcome measures (PROMs)

- Initially, the OECD should seek to standardise and harmonise PROMs benchmarking for patients who have undergone hip and knee surgery, as these currently represent the widest application PROMs in OECD countries. The OECD should compare the PROMs being used for these conditions, and explore options to standardise and/or harmonise them. Generic PROMs (such as EQ-5D) should be used alongside condition-specific PROMs, since the combination of the two will provide a fuller picture of patient outcomes.

- The OECD should extend its PROMs work to longitudinal studies of chronic disease patients, as this is where the need for more information with regards to care co-ordination is most urgent. This work should begin with cancer, as this is an area where several PROMs already exist, and where the OECD already collects complementary data on survival. Work should later be extended to emergency care, mental health care (including dementia), long-term care, palliative care, informal care, and preventive care.

- The OECD should also develop PROMs for conditions, sectors and services where instruments are currently lacking. Of particular importance are patients with multiple, chronic conditions. In this group, a combination of disease-specific PROMs, generic PROMs as well as PREMs will be essential to fully understand the performance of health care systems.

4. Providing technical assistance to embed patient-reported indicators into clinical practice and into health care information systems

- The OECD should work with patient and professional groups to identify the best methods to embed the collection, analysis and use of patient-reported indicators into routine clinical work.

- The OECD should support learning between countries on embedding patient-reported indicators into electronic health records, clinical registries, mobile apps and other sources.

- An initial step would be to document the different methods already used to embed PROMs and PREMs within the clinical practices and information infrastructure of different health systems, barriers to adoption and use, as well as solutions to overcome those barriers.

Note

1. Disease level: collection and analysis of patient-reported experience/outcomes in particular patient groups, e.g. those suffering from dementia, specific chronic conditions, as well as patients with multiple chronic conditions or hip fracture.

   - Sector level: collection and analysis of patient-reported experience/outcomes in particular health-care sectors, e.g. clinical care, long-term care and mental health care.

   - Service level: collection and analysis of patient-reported experience/outcomes in individual hospitals, clinics or other facilities, to make comparisons at a national level.

   - System level: collection and analysis of patient-reported experience/outcomes at a national or system level, e.g. by working with countries to encourage measures across the entire health system and across the full pathway of patients' care.

Introduction

The OECD has unrivalled international experience in the collection, analysis and dissemination of health system metrics. Its well-developed repository of health statistics has been regularly refined since it was launched in the 1980s. The *Health* Statistics *Database* and high-profile biennial report, *Health at a Glance*, are popular tools widely used to conduct international appraisals of health systems and population health status. The genesis of this work was the collection of comparable data on health expenditure. It has evolved over time to encompass other valuable data on health system performance, such as the quality of health care.

To date, however, OECD data have largely focused on health system inputs, activities and costs. There are much fewer measures of outcomes, particularly those directly reported by patients. This leaves governments with only a partial view of how well their health systems are responding to patients' needs. If countries are to be well-equipped to meet the challenges presented by ageing populations and the accompanying rise in chronic disease and multiple morbidities, it is essential that data collected are relevant and actionable, and correspond to what matters most to patients.

Areas where need is complex and growing are particularly poorly measured. For example, much more is known about elective procedures routinely performed in hospitals than the effective management of chronic disease in primary health care. This is an information gap requiring prompt attention. Of particular urgency is the need for more information on the co-ordination of care, primarily for patients with chronic disease and long-term conditions. The public health burden these conditions will pose in coming years demands stronger scrutiny of how effectively health systems are providing integrated care, to minimise the risk of medical errors and other outcomes that are unacceptable to patients as well as costly for health systems.

"Patient-centred care" is an objective that is regularly used by policy makers and clinical leaders as defining the way in which they believe

health care should develop. However, metrics on whether or not this is being delivered are largely absent, even if there is wide consensus that there are huge benefits to giving providers as well as decision makers such information.

Information deficits also exist in the outcomes of patients in mental health care, emergency care, long-term care, palliative care and preventive care, as well as the outcomes of informal carers. If governments are to improve the quality of life and the outcomes of more vulnerable patients, there will be a need for more – and better – information in these services.

There is great promise in translating these metrics into actions that can drive improvements in patient experience and outcomes. There are potential benefits for policy makers, health care providers at both an organisational and individual level, and for patients. Hospital-level comparisons, for example, can identify wide variations in practice and the overuse and misuse of treatment, thus providing opportunities to minimise the wasteful use of resources.

Information reported directly by patients can offer insights that cannot be identified through other means. For example, the only way to know whether patients recovering from prostate cancer experience problems with incontinence and erectile dysfunction is to ask them. Similarly, patients who have multiple chronic conditions are more at risk of experiencing un-coordinated care with greater probability of complications or errors as they move across the care pathway. This information would enable patients to compare their experience and outcomes relative to other patients.

For governments, information on patient outcomes would support policy decisions about how to make health systems more centred around the needs of patients, and more efficient in addressing those needs. Comparing the performance of their health system with that of other countries in this area will also help identify weaknesses and trigger a compelling case for change.

The political consequences of poor health system performance for citizens and voters offer governments a powerful incentive to collect more intelligence on patient outcomes to drive improvements. There is too high a cost associated with not collecting information on the experience and outcome of patients' care, in the form of missed opportunities to improve clinical practice and patient quality of life. Additionally, early evidence suggests that the use of PROMs to analyse

and inform decisions in health care has the potential to improve patient outcomes while in parallel reducing costs (Basser, 2015).

Enough progress has been made to demonstrate the potential utility of such approaches. But – with a few notable exceptions – the development of measures in this area has been undertaken in a fragmented way on a national basis, and there is very little international measurement in this area. In the absence of international co-ordination, the risk that each country develops its own standards that are not comparable across countries is high. This would deny policy makers the ability to benchmark their outcomes with other countries. In addition, in the absence of strong leadership at the political level, there is a risk that measures focus on areas where measurement is most practical – generally hospitals and elective surgery – rather than where measurement is most needed – such as in the primary care, mental health care, emergency care, long-term care and palliative care sectors.

The OECD has the political impetus to bring about this much-needed change. In the health domain, the OECD has long demonstrated its leadership in the collection, dissemination and analysis of internationally comparable health statistics. The OECD is thus in a strong position to implement standardised, validated instruments for the collection of patient-reported indicators data across its member countries and beyond.

1. The importance of patient-reported indicators of health system performance

Patient-reported indicators of health system performance largely relate to patient-reported experience measures (PREMs, such as whether the patient feels they were adequately involved in important decisions about their care), and patient-reported outcome measures (PROMs, such as whether the patient is free of pain after an operation care). Some OECD countries are conducting PREMs surveys and, to a lesser extent, are experimenting with PROMs. The OECD already has some comparable indicators on patient experience, developed as part of its regular data collection on quality indicators.

The statistics routinely collected, however, provide an insufficient picture of the outcomes of health care. Furthermore, countries are overwhelmingly "doing their own thing", presenting little opportunity for cross-country comparative analysis. This section provides a summary of how patient-reported indicators are being developed and used in OECD countries, and how the collection of more of these metrics would bring benefits to policy makers, patients and clinicians.

1.1. The use of patient-reported experience measures (PREMs) in OECD health systems

Measuring and monitoring patient experience can inform changes to clinical practice that are necessary to improve quality of care. Factoring the patient voice into health system design can also help in the provision of health care that is more responsive to patients' needs. In the case of chronic disease, in particular, the growing emphasis on patient self-management will make capturing metrics on patient experience even more important.

PREMs: Measure patients' perceptions of their experience of care by focusing on the process of care and how that has an impact on their experience. Examples: Did the patient wait long for treatment? Did the patient feel they were involved in decision making?

While much PREMs activity is occurring at a national level, however, there are few instruments facilitating international comparisons. This lack of information is a wasted opportunity for policy makers to understand failures that have led to unsatisfactory experiences in care.

A positive patient experience should be considered an outcome in its own right, and correlates well with other measures of quality

There is some evidence of the relationship between patient experience with the process of care and outcomes. One study found that – after adjustment for post-discharge health status and other clinical factors – patients experiencing worse hospital care had lower ratings of overall health, physical health and were more likely to have chest pain one year after an acute myocardial infarction (AMI) than other patients. However, the association between a negative hospital experience and subsequent chest pain may be offset by more positive outpatient experiences. The most frequent problems occurred in information and education, emotional support, involvement of family and friends, and continuity and transition to home. The study authors cite the quality of communication with patients as a key factor. Patients experiencing difficulty obtaining clear guidance about their condition and treatment may be less likely to take their medications appropriately after discharge, make lifestyle changes that would improve their recovery, and may be less likely to attend follow-up outpatient appointments or to report concerns. Such patients also may be at greater risk of anxiety or depression, which are associated with worse outcomes after an AMI (Fremont et al., 2001).

A study of women with breast cancer showed an association between ongoing cancer therapy with tamoxifen four years after the initiation of treatment, and patient-centred care. The proportion of patients with ongoing tamoxifen use was lower for those reporting less support than needed, a less-than-adequate role in decision making regarding tamoxifen use, inadequate input of a doctor in making decisions about tamoxifen use, and not being told in advance about the medication's side effects. This was the case after adjusting for the severity of side effects, and other demographic and clinical factors (Kahn et al., 2007). Another study found a positive, albeit modest, correlation between measures of patient experience with process measures of clinical quality in prevention and disease management in primary care. However, there

were no significant correlations between patient experiences of care and clinical outcomes (Sequist et al., 2008).

OECD health systems are increasingly applying PREMs as a critical indicator of performance

Many PREMs initiatives have been undertaken in previous decades, often modelled on the work of the Picker Institute in the United Kingdom and the Consumer Assessment of Healthcare Providers and Systems (CAHPS) initiative in the United States. The Picker Institute developed the Principles of Patient Centred Care in 1987, which became a framework used internationally to support high-quality patient-centred care. In 2002, it designed and established England's first National Health Service (NHS) national survey programme for patient experience. The CAHPS initiative was launched in 1995 by the Agency for Healthcare Research and Quality, with standardised questionnaires measuring the patient experience.

At a national level, in almost all OECD countries PREMs are collected through surveys, covering population samples of patients who experience inpatient or outpatient care. Some countries are developing tools to evaluate patient experience with specific care needs to improve the delivery of care for particular population groups. These may be condition-specific surveys (e.g. cancer and diabetes) or care-specific (e.g. maternity or psychiatric care). For example, Norway conducts surveys focusing on people with specific illnesses, including adult and paediatric patients who received mental health care. In the Netherlands, PREMs are collected from people with diabetes, asthma, heart failure and cancer, covering providers such as general practitioners, physiotherapists, hospitals and nursing homes (Fujisawa and Klazinga, forthcoming). There are many conditions for which PREMs are not routinely collected. For example, the experience of patients with rheumatoid arthritis is not routinely measured in many countries. A PREM for rheumatoid arthritis has recently been developed, piloted and validated, and is being used in a National Clinical Audit in England and Wales (Bosworth et al., 2015). There are also no PREMs that allow an assessment of patient experience with multiple chronic conditions.

Box 1.1. The use of PREMs in France

France's e-Satis initiative measures patient satisfaction and experience in hospitals. Importantly, the survey also includes questions about care co-ordination, including questions about hospital discharge and how well care is co-ordinated between hospitals and GPs.

The information is fed back to hospitals to help them improve quality. It also provides information and choice to the public. The data are also used for the purpose of pay for performance, in that hospitals receive bonuses for good results. There are regional-level data and national level data.

The e-Satis data was published for the first time in 2016. If hospitals do not get satisfactory results then a note is made on their accreditation record, which is public. Accreditation is compulsory for all public and private hospitals in France.

Service-level patient experience measures are being used to inform health care regulators for inspection, regulation and accreditation. For example, the Czech Republic awards "Satisfied Patient" certificates to the health providers with outstanding performance related to patient experiences (Fujisawa and Klazinga, forthcoming).

The co-ordination of care across the patient pathway is critical in reducing the risk of safety mishaps caused by failures in communication between providers. This is an area where there would be great benefit in learning more about the patient experience, but knowledge is limited. In a US study, a self-report measure of the quality of care transitions capturing the patient's perspective was developed for adult patients discharged from hospital. The Care Transitions Measure (CTM) was found to have high internal consistency and reliability. It was developed with the input of focus groups to cover the domains of information transfer, patient and caregiver preparation, support for self-management, and empowerment to assert preferences. It was found to have the capacity to discriminate between patients who had an emergency department visit or rehospitalisation for their index condition, and those who did not, and to converge with patients' reports of negative experiences after their discharge from the hospital (Coleman et al., 2005). The CTM has been validated in other populations (Bakshi et al., 2012; Parry et al., 2008).

Public and private payers in the United States and some other countries are recognising patient experience as a quality component, and factoring the results of patient experience surveys into provider payment structures. In 2006, the Medicare Modernisation Act tied hospitals'

Medicare payment rates to collecting and publicly reporting hospital CAHPS data. Most US hospitals collect and publicly report on standardised core CAHPS survey questions (Aligning Forces for Quality, 2010).

Opportunities for comparing PREMs across countries are currently limited

Despite ongoing PREMs initiatives taking place at a country level, there is still little opportunity to compare the experience of patients at an international level. Nineteen OECD countries provide data on the patient experience with ambulatory care, and the data are reported in the OECD Health at a Glance as an indicator of quality of care. This includes:

- doctor spending enough time with patients during a consultation;
- doctor providing easy-to-understand explanations;
- doctor providing the opportunity to ask questions and express concerns; and
- doctor involving patients in decision making affecting their care and treatment.

However, the scope of these indicators is limited to the outpatient sector. The OECD does not report on patient experience in inpatient care, mental health care, emergency care, long-term care or palliative care. Nor does the OECD report on patient experience for specific conditions or for patients with multiple chronic care needs.

Eleven countries participate in the Commonwealth Fund's International Health Policy Surveys (Box 1.2). Unlike the OECD's core set of patient experience questions of the general population, the Commonwealth Fund surveys cover both general population and more specific population groups. Another key difference is the Commonwealth Fund surveys extend to access and use of emergency departments, waiting times to see physicians, gaps in care co-ordination, and cost as a barrier to health care.

Box 1.2. Commonwealth Fund International Surveys

The Commonwealth Fund conducts international surveys of the general population, and also surveys targeting more specific population groups. Eleven countries participate in the survey: Australia, Canada, France, Germany, the Netherlands, New Zealand, Norway, Sweden, Switzerland, the United Kingdom and the United States.

The general population survey consists of computer-assisted telephone interviews of random samples of adults, using a common questionnaire translated and adjusted for country-specific wording as needed. In the last survey, conducted in 2013, response rates ranged from 11% in Germany and Norway to 33% in Switzerland, introducing the possibility of bias. Questions include the extent to which participants skipped health care because of cost, waiting times to see physicians, whether patients have email access to doctors, out-of-hours access and emergency department use, administrative costs and complexity of health insurance, and whether participants believe their country's health system needs major change.

Two surveys are more population-specific. One survey focuses on older adults aged 65 and over, and uses computer-assisted telephone interviews to ask about access to care (related to cost, access to out-of-hours primary care, and avoidable use of emergency departments), the existence and management of chronic conditions, patient experience with care co-ordination (e.g. lack of communication between providers), patient engagement, social care needs, and end-of-life care planning (patients have discussed their care wishes with their doctor or family or have a written plan in place).

Another survey focuses on adults aged 18 and over with complex care needs who meet at least one of four criteria: they rate their health as fair or poor; report receiving medical care for serious chronic illness, injury, or disability in the past year; or had surgery or had been hospitalised in the previous two years. This survey covers the experience of patients with a medical home, care co-ordination, medical errors, patient satisfaction, and cost and access problems.

Source: The Commonwealth Fund, http://www.commonwealthfund.org.

The PREMs work to date carried out by countries at a national level, and by the OECD and the Commonwealth Fund at an international level, is a good basis to build upon. A structured review of national and cross-national surveys of patient experience for OECD and non-OECD European Union member countries undertaken since 1997 found that despite the existence of several instruments, few had been trialled for cross-country use. Among them is the World Health Organization's Health System Responsiveness study of 60 countries, conducted in 2000-01, and World Health Survey from 2002 which also included questions on responsiveness (Garratt et al., 2008).

Building on the more recent work of the OECD and the Commonwealth Fund, by collecting more indicators across more sectors

and as patients move across the care pathways, would provide an opportunity for cross-country comparisons that would be useful for policy makers, and help drive improvements in health system performance.

There is also potential for the OECD to collaborate with commercial research companies that capture the patient experience. One such company, Ipsos MORI, conducts research for the NHS in England as part of its health and social care work. In a recent example, Ipsos MORI researchers were involved in the BRIGHTLIGHT Survey, a PREM for young people with cancer. Ipsos MORI moderators and researchers facilitated focus groups and conducted telephone interviews (Taylor et al., 2015).

1.2. The use of patient-reported outcome measures (PROMs) in OECD health systems

A growing number of countries are showing an interest in using patient-reported outcome measures (PROMs) to learn more about patient outcomes. The use of patient-reported indicators alongside other quality and outcomes metrics provides a more comprehensive picture of health system performance that can benefit not only patients, but also individual clinicians, health care providers and policy makers. The collection of this information at an international level is critical, therefore, to delivering health services shaped around the needs of patients.

PROMs: Measure patients' perceptions of their health status, clinical outcomes, mobility and quality of life. Examples: What was a patient's mobility like before a hip replacement, and did it improve after the intervention? Does a patient's condition limit their ability to do strenuous activities such as jogging, skiing or cycling?

However, opportunities for cross-country comparisons are even more limited than PREMs. This gives policy makers only a superficial understanding of the outcomes of health care.

PROMs have the potential to drive improvements in clinical practice

PROMs instruments can be disease-specific or more generic. Those that are disease-specific focus on the symptoms and impact on function of a specific condition, such as whether a patient can do the shopping independently after an intervention like knee surgery. An example of a generic PROM is the EQ-5D, developed by the EuroQol Group. Patients can report, for example, extreme pain or discomfort, or anxiety. The use of disease-specific and generic tools together can provide a fuller picture of a patient's outcomes.

Box 1.3. Generic PROMs relating to quality of life

EQ-5D

EQ-5D is a generic health outcome measure, and is applicable to a wide range of health conditions and treatments by identifying 243 possible health states. Patients describe their own health state on five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression, and one of three levels of severity is chosen for each dimension: no problems, some/ moderate problems or extreme problems.

The EQ-5D generates the EQ-5D Index which is a health profile that can be made into a global health index with a weighted total value for health related quality of life, representing the patient's description of his own health and how this health state is perceived by the general population. There are several different value sets to calculate the utility scores, and each value set represents the preferences of the population from which it was derived. Thus, comparisons of results using utility indices calculated with different value sets may be difficult.

The original EQ-5D, which has three levels of response options (EQ-5D-3L), is most commonly used and has been validated for patients with osteoarthritis. The EQ-5D-5L is an extended version of the EQ-5D that has five response options for each dimension, The EQ-5D-5L has better psychometric properties (such as better responsiveness and lower ceiling effects) than the EQ-5D-3L, and increased use of the EQ-5D-5L is anticipated in clinical studies and registries.

Short Form 36 (SF-36)

The Short Form 36 health survey (SF-36) includes eight dimensions of health: physical functioning, bodily pain, role limitations due to physical health problems, role limitations due to personal or emotional problems, emotional well-being, social functioning, energy/fatigue, and general health perceptions. It also includes a single item that provides an indication of perceived change in health.

The eight scaled scores are the weighted sums of the questions in their section. Each scale is directly transformed into a 0-100 score. A score of zero is equivalent to maximum disability and a score of 100 is equivalent to no disability. The scores of the eight scales are summarised into one physical and one mental scale component. The SF-36 is the most commonly used generic PROM in clinical trials, and is psychometrically sound for patients who have osteoarthritis.

SF-12

SF-12 is a generic health outcome measure, which consists of 12 items derived from the 36-item score, SF-36. The SF-12 gives two summary scores; Physical Component Summary (PCS) and Mental Component Summary (MCS), ranging from 0 to 100 with higher scores being better. For routine follow-up in joint replacement registries the SF-12 is considered as preferred instrument to the SF-36.

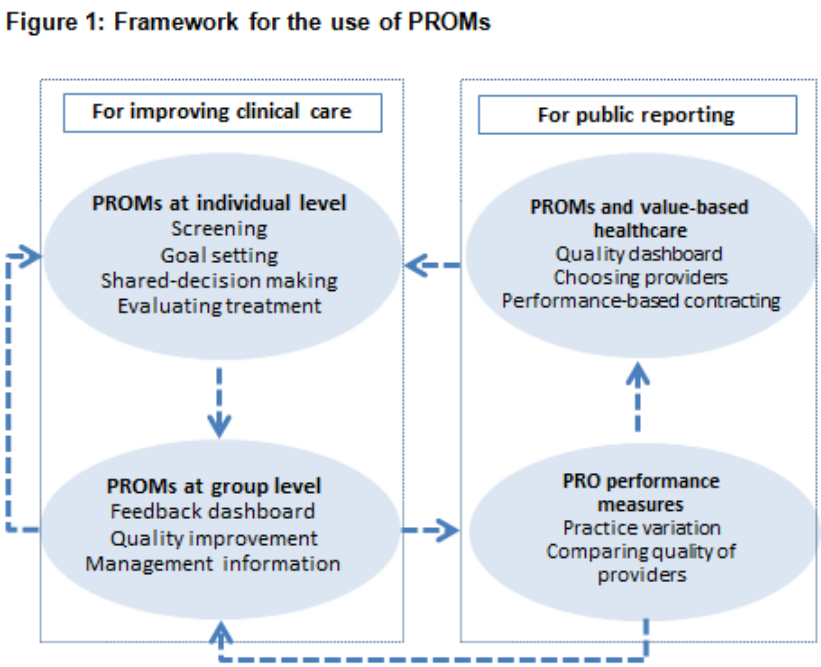
SF tools require licensing, but the equivalent Veterans Rand 12-item survey (VR-12) and 36-item survey (VR-36) are available in the public domain and free of charge.

Crosswalk algorithms

A recent study successfully evaluated a probabilistic mapping approach on EQ-5D utility scores based on SF-12 responses using Bayesian networks. Such a mapping approach allows for crosswalk algorithms to convert SF-12 responses to EQ-5D index scores, enabling comparisons between the tools.

Figure 1.1 presents the use of PROMs in clinical practice, quality improvement and performance measurement. The figure visualises the use of PROMs for internal use by provider organisations in health care to improve clinical care; and for external use by stakeholders in performance measurement and public reporting.

Figure 1.1. Framework for the use of PROMs



Source: Philip van der Wees, Radboud University Nijmegen.

To evaluate the evidence for the use of PROMs to drive quality in healthcare, the Secretariat conducted a rapid review of reviews, and identified recently published individual studies published after the most recent review. Thirteen reviews were identified, including twelve quantitative reviews and one qualitative review. Annex A shows the characteristics and results of these systematic reviews. In summary, the reviews show that the feedback of PRO data to healthcare professionals in clinical practice can lead to improvements in the quality of patient care, with a stronger evidence base for improvements in the care process than health outcomes. In oncology care a strong evidence base exists for

the use of PROMs to detect symptoms, to improve communication between clinicians and patients, and for patient satisfaction. In mental healthcare insufficient evidence was found to support the routine use of PROMs to improve clinical practice.

PROMs are being used in clinical practice to help assess patients' symptoms, function and health-related quality of life, and respond with appropriate treatment. They are also being used to determine the effectiveness of treatments. For example, in one study, focus groups of patients who survived stroke identified as the most relevant and meaningful outcomes being alive at home, without recurrent stroke, and without being hospitalised for complications. The study examined the association between warfarin treatment and the outcomes in patients after ischemic stroke with atrial fibrillation, compared with patients given no oral anticoagulant at discharge. Patients treated with warfarin at discharge had a significantly lower risk of major adverse cardiovascular event over two years and were more likely to spend more days alive and out of hospital than those not receiving any oral anticoagulants at discharge (Xian et al., 2015). The study is an example of how patient-centred research can be used to support decision making by patients and clinicians.

PROMs can also help improve communication between patients and physicians. In one study of cancer patients, those who completed PROMs discussed more symptoms during consultations with oncologists, particularly pain, fatigue and nausea and vomiting. However, PROM feedback had no impact on discussion of patients' functioning (Takeuchi et al., 2011). Another study investigating clinical paediatric rheumatology care monitored health-related quality of life using electronic PROMs. It found use of the PROM increased discussion of psychosocial topics as well as the paediatric rheumatologist's satisfaction with the care provided during the consultation. Parents and children also evaluated the tool positively (Haverman et al., 2013). In a study of schizophrenia patients, measuring quality of life had a positive impact on patient satisfaction, but no effect on improving clinical outcomes. The latter suggests that clinicians did not optimally use the data (Boyer et al., 2013).

At a national level, the collection and reporting of patient-reported data can inform policy makers' decision making in terms of resource allocation, identify variations and inequalities, and assist them in being more responsive to patients' needs. Additionally, comparing the outcomes of patients using different health services at a national level

can identify outliers requiring attention. Publicly reporting performance metrics also provides the public with important information to assist them in making choices about treatments and hospitals. This transparency is fundamental to a health system that is truly patient-centred, although education campaigns will be needed to generate public awareness of the existence of this information and its purpose. Further research will also be required to assess the barriers to achieving these goals, so that the collection and reporting of these metrics translate into actions that drive improvements.

At an organisational level, there are is a clear potential from the use of PROMS. For example, Bupa Hospitals began collecting PROMs data in 1998, originally inspired by a desire to identify clinical "bad apples". However, it also recognised that PROMs had the potential to drive continuous quality improvement, change clinical practice, and provide feedback to clinicians and patients. Additionally, by publishing hospital-level PROMs results on their websites, Bupa Hospitals were able to promote the health-related quality of life benefits of the interventions they provided (Devlin et al., 2010).

Bupa focused on higher-volume and higher-risk procedures to ensure the results would be statistically relevant. The PROMs data were used as an indication of where other process and outcomes data might require deeper analysis. In one example of how Bupa used the PROMs data to promote best practice, consistently higher than average health gain following hip replacement was identified at one hospital. An investigation revealed that the hospital's physiotherapy department had started an intensive pre-operative work-up of patients planning to undergo hip replacement surgery, meaning they were better prepared both for their procedure and recovery periods (Devlin et al., 2010).

While the use of PROMs to improve clinical practice is still relatively new, there are nevertheless several examples of their use in this regard. ICHOM is aware of about 185 organisations implementing one ICHOM Standard Set, 33 organisations implementing more than one and 14 disease registries measuring at least one. ICHOM is directly supporting 40 organisations through standard set implementation. The Aneurin Bevan Health Board in Wales, for example, has recently implemented the Parkinson's Disease Standard Set and is preparing to implement the Low Back Pain and the Cataracts Standard Sets (ICHOM, 2015a).

The examples from the published scientific literature cited in Section 1.2 suggest that the collection and reporting of PROMs data has the potential to improve clinical practice and the process of care. Ideally, health care providers at an organisational and individual level should use the information to identify where they need to improve, why they may not do as well as other providers, and what strategies they can adopt to achieve improvements. However, the evidence on the impact of PROMs on patient outcomes remains limited.

A review of the use of PROMs in cancer clinical practice also suggests that more research is needed to translate the knowledge gained from PROMs into actions that change behaviour to achieve an effect on health outcomes (Howell et al., 2015).

Box 1.4. The use of clinical registers in selected OECD health systems

Sweden has numerous national Quality Registries (NQRs) of which 93 include some form of PROM or PREM measure (Nilsson, 2016). Half of these include some type of generic measure, and more than half include disease-specific PROMs. Of the generic measures the most common measures are the EQ-5D and the SF-36/RAND-36. The main use of PROMs in Sweden has been in clinical trials or other research. The use of PROMs in clinical practice and local quality improvement is increasing. Examples of NQRs including PROMs are the Hip Arthroplasty Registry (using the EQ-5D), the Cataract Registry (using the Castquest-9SF), and the Stroke Registry (using EQ-5D and other measures).

A Framework for Australian Clinical Quality Registries was developed to stimulate the number of data collections that capture and report process and outcomes data for specific clinical conditions or interventions. The development of a number of high-priority national registries has the potential to address the current gap in health care quality measurement and inform improvements in the quality of patient care. The Framework, endorsed by the Australian Health Ministers' Advisory Council (AHMAC) in March 2014, describes national arrangements for clinical quality registries (ACSQHC, 2014). In Australia several examples exist of national clinical outcomes registries – sometimes in collaboration with New Zealand – such as the Prostate Cancer Outcomes Registry (PCOR-ANZ). Data collection includes a quality of life questionnaire (EPIC-26) at 12 and 24 months post active treatment, from a follow-up phone call to the patient (Nag, 2016).

In the Netherlands multiple clinical registries exist which have started to include the use of PROMs. The Dutch Institute for Clinical Audit (DICA) operates 23 national registries, of which nine registries include PROMs (DICA, 2016). Examples of registries including PROMs are the Dutch Surgical Spine Registry (e.g. including NDI, ODI) and Bariatric Surgery (RAND36). DICA is moving towards stimulating quality improvement based on data feedback.

The impact of PROMs feedback on outcomes of care is limited, however, underlining the importance of further activity and research in this area. A study in orthopaedic surgery by Boyce and Browne (2015)

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

Rigorous studies of the effects of using PROMs as a performance measurement tool are relatively scarce reflecting the limited number of programmes to date, and the still-nascent evidence base for patient-reported performance indicators. A systematic review by Boyce and Browne (2013) identified only one study of performance feedback at the group level and it found no effect on performance. Chen and colleagues (2013) found no studies that evaluated the use of PROMs in oncology setting for quality improvement, transparency, accountability, public reporting or system performance.

A recently published study by Varagunam (2014) suggested that hospital performance was not altered by introduction of routine patient reported outcome measures in surgery in England. The authors concluded that the manner in which results are communicated, the need for timely feedback, and inclusion of suggested action steps to improve PROMs might be necessary. Partridge (2016) conducted a non-controlled quality improvement in total knee replacement using PROMs data. Statistically significant differences in outcomes after surgery when using different brands of implant – measured by the Oxford Knee Score – was reason for changing to the better performing implant.

Qualitative studies show that the use of PROMs as a quality improvement tool is complex, and tailored feedback to support interpretation of PROMs is important to stimulate quality improvement. This becomes even more important with the feedback of aggregate data in understanding variation in outcomes between clinicians or provider organisations (Boyce et al., 2014; Howell et al., 2015). These findings echo studies of the use of patient experience data in performance measurement (Friedberg et al., 2011). These studies suggest the need to embed performance measurement in a formal quality improvement programme.

Further knowledge of how PROMs are deployed by health system should emerge from two literature reviews currently underway. A realist review by Joanne Greenhalgh and colleagues is aimed at understanding by what means and in what circumstances the feedback of PROMs data leads to the intended service improvements (Greenhalgh, 2014).

Concalves and colleagues are currently conducting a Cochrane Review with the objective to assess the impact of the routine use of PROMs in clinical practice on the process of care; patients' and professionals' experiences of care; and health outcomes (Concalves, 2015).The results of the Cochrane review are expected in the spring of 2017.

At country-level, PROMs are rarely measured regularly and systematically

The collection of PROMs at a system-wide level is in its infancy

The NHS in England was the first health system in the world to introduce the routine collection of PROMs data at the system level. Since 2009, the Department of Health has required the routine measurement of PROMs for all NHS patients in England before and after receiving surgery in the case of four elective procedures. Patients complete both generic and condition-specific surveys for knee replacement surgery, hip replacement surgery and varicose vein surgery. For hernia repair, they complete only a generic survey (Table 1.1).

Table 1.1. PROMs programme in the NHS England

Treatment	Condition-specific PROM	Generic PROM
Knee replacement	Oxford Knee Score	EQ-5D (including EQ VAS)
Hip replacement	Oxford Hip Score	EQ-5D (including EQ VAS)
Varicose vein removal	Aberdeen Varicose Vein Questionnaire	EQ-5D (including EQ VAS)
Hernia repair	No instrument	EQ-5D (including EQ VAS)

Source: Health & Social Care Information Centre (2015). Note: EQ VAS = EQ Visual Analogue Scale.

The requirement to collect PROMs for the four procedures is part of the NHS contract for acute services, with the aim of improving clinical quality and patient outcomes. It is part of the NHS Outcomes Framework, designed to provide accountability for the outcomes the NHS delivers, and monitor the performance of health services. The information is publicly reported on the My NHS website, enabling hospital-level comparisons.

While England is the most advanced in implementing a system-wide PROMs programme, a number of other countries are showing an interest in collecting PROMs data in sectors such as elective surgery, mental health care, long-term care, palliative care, and informal care (Box 1.5).

Box 1.5. The use of PROMs in specific sectors

Elective surgery

Sweden and the Netherlands each have Hip Arthroplasty Registers that collect PROMs data for total hip replacement patients. The New Zealand Joint Registry collects PROMs data from a random sample of 20% of joint replacement surgery patients. Various PROMs initiatives also exist in different provinces of Canada. For example, Alberta has the Edmonton Heart and Lung Transplant Clinic pilot project. Pre-operative health information is collected when a patient is placed on the transplant list, and post-operative information is collected each time the patient attends the clinic (CIHI, 2015). Sweden also has a National Spine Register, *Swespine*, as a basis for quality assurance and improvement. It measures health-related quality of life and spine-related disability, with follow-up assessments one, two and five years after surgery (Strömqvist et al., 2013).

Mental health care

In the Netherlands, routine outcome monitoring (ROM) is used to improve the quality of mental health care. ROM is the systematic measurement of treatment outcomes in routine clinical practice, and can be used by patients and clinicians to monitor treatment progress. It can help determine psychiatric diagnosis, symptoms and psychosocial functioning in every phase of treatment. Anonymised ROM data can be used to conduct epidemiological research, and for benchmarking purposes. The Hamilton Depression Rating Scale and Montgomery-Åsberg Depression Rating Scale are examples of disorder-specific rating scales that measure symptom severity in major depression (van Noorden et al., 2013).

Long-term care

InterRAI is a not-for-profit network of researchers in more than 30 countries. It aims to promote evidence-based clinical practice and policy through the collection and interpretation of data about the characteristics and outcomes of people in a range of settings. Assessment instruments have been developed for a range of settings, including long-term care facilities. To a lesser extent, InterRAI has been used in home-care settings. InterRAI risk-adjusted quality indicators for nursing homes include mobility, behaviour, bladder continence, mood and pain (Carpenter and Hirdes, 2013). InterRAI instruments are administered by assessors who make observations based on interviews with patients, as well as consulting caregivers and staff, and reviewing all available records. For example, to determine whether an elderly woman prepares her own meals or performs her own housekeeping in a home-care setting, the assessor will consult the person, her caregiver, community service providers, and any available records. Additionally, a number of items are specifically addressed to the person being assessed, such as aspects of mood and self-reported perception of health (Gray et al., 2009). Among InterRAI's survey instruments are a series of site-specific Subjective Quality of Life (QoL) instruments, designed to give people enrolled in formal care programmes the opportunity to voice their perceptions in domains such as relationships, environment, comfort, food and participation in meaningful activities. They are intended to be used during interviews with patients who have the cognitive capacity to respond, or can be completed by the person and returned by mail (InterRAI, 2015).

There also exist disease-specific questionnaires. These include the Parkinson's Disease Questionnaire (PDQ-39). It covers the dimensions of mobility, daily living activities, emotional wellbeing, stigma, social support, cognitions, communication, and bodily discomfort (Jenkinson et al., 1997). For patients with dementia, the DEMQOL instrument and DEMQOL-Proxy have been used for patients and their carers respectively. The DEMQOL and DEMQOL-Proxy have been

found to provide a method to evaluate health-related quality of life, and are recommended to be used together in the case of mild and moderate dementia. For patients with severe dementia, only the DEMQOL-Proxy is recommended (Smith et al., 2005).

Palliative care

The Palliative Care Outcome Scale (POS) was developed for patients with advanced cancer. It consists of two almost identical measures, one completed by staff, the other by patients. It assesses a range of domains including pain, patient and family anxiety, support, information, feeling that life is worthwhile, feeling of self-worth, time wasted on appointments relating to health care (e.g., waiting for transport or repeating tests), whether practical financial or personal matters resulting from illness have been addressed, and the patient's main problems in the previous three days (Hearn and Higginson, 1999). The POS has been used in countries including Germany and Austria (Bausewein et al., 2005).

Informal care

The carers of patients with long-term conditions are at risk of high levels of stress, depression and anxiety, making it important to assess their health-related quality of life. The Short-Form Health Survey (SF-36) has been used to assess the quality of life of carers in several countries with different population groups, and has been translated and validated in many languages. A British study of carers of patients with motor neuron disease, multiple sclerosis and Parkinson's disease used the SF-12 (derived from the SF-36), as well as the Carer Strain Index and a newly-developed questionnaire on health and social care experiences. The study confirmed carer wellbeing was compromised (Peters et al., 2013b). In a study in Spain, researchers attempted to use the SF-36 in carers of patients with dementia. The study found 37.6% of the female carers indicated that their health was worse or much worse than the previous year, compared with 26.2% of females in the control group. In males, the differences were not statistically significant (21.6% of the male carers compared with 20.7% of males in the control group). However, the authors noted that a limitation of the study was that the SF-36 had not been validated with respect to carers of patients with dementia (Argimon et al., 2004). A limitation of both studies is that they are cross-sectional surveys.

State-wide performance measurement for several conditions

Minnesota Community measurement is a state-wide initiative in the United States for performance measurement. State- wide and medical group rates of performance on quality measures are published for patients in Minnesota Health Care Programs. The annual Health Quality Report includes 32 measures for different conditions. PROMs are included for several conditions such as spinal surgery (average change in the Oswestry Disability Index (ODI) at three months post-operative), total knee replacement (average Oxford Knee Score (OKS) change at one year post-operative), and depression [percentage remission at six months based on the Patient Health Questionnaire (PHQ-9)] (Snowden, 2015).

In a survey and set of structured interviews[1] commissioned by the HLRG, national experts confirmed that the use of patient-reported experience measures (PREMs) is more advanced than the use of PROMs. PREMs have been used for many years to measure and improve quality of care. The use of PROMs for purposes other than research is relatively new. In several countries (Australia, Canada, Netherlands) PROMs questions have been added to PREMs surveys, preceding or in

conjunction with the implementation of PROMs programmes. This is the case in the Australian state of New South Wales, for example, which has added some PROMs questions onto the PREM for hospitals.

National experts also emphasised that PROMs and PREMs serve both distinct and common objectives. Differences in how PREMs and PROMs are administered (e.g., timing, frequency) should be considered when evaluating opportunities and the feasibility of administering these together. PROMs are used mostly for research and to facilitate shared decision making between clinicians and patients to improve clinical practice, and sometimes provide other purposes for quality improvement and public reporting. Interviewees confirmed that the use of PROMs instruments have potential to improve process and outcomes of care, and to reduce inappropriate care. However, the use of PROMs to assist decision making about resource allocation is not common yet.

Interviewees confirmed that the most common used purpose of PROMs data collection was the use by clinicians and patients to stimulate shared decision making and improve outcomes. Elective surgery, mental health, and cancer are the most common diseases/conditions for which PROMs are currently collected. PROMs were sometimes used in long-term care, palliative care and preventive care. In informal care PROMs use was not reported by any of the 13 participants. Significantly, interviewees also confirmed that there is inconsistency in the selection of PROMs both between and within countries. In some countries professional bodies of (regional) governmental bodies mandate the use of specific PROMs.

Concerning elective surgery, the most frequently mentioned procedures were knee surgery (n=9 countries), hip surgery (n=8 countries) and cataract surgery (n=4 countries). Commonly used PROMs for hip surgery are the Hip Injury and Osteoarthritis Outcome Score – Physical Function Shortform (HOOS-PS) and the Oxford Hip Score (OHS). Commonly used PROMs for knee surgery are the Knee Injury and Osteoarthritis Outcome Score – Physical Function Shortform (KOOS-PS) and the Oxford Knee Score (OKS). Commonly used disease-specific PROMs for knee surgery are the KOOS-PS and the OKS. Commonly used disease-specific PROMs for hip surgery are the HOOS-PS and the OHS (see Annex B). In cataract surgery all four countries use the Catquest 9-SF.

In cancer care two main PROMs were identified in prostate cancer: the European Organization for Research and Treatment of Cancer,

Quality of Life Questionnaire-C30 (EORTC-QLQ-C30); and the Expanded Prostate Cancer Index Composite (EPIC-26/16). COPD was the most common chronic disease for which PROM are used, with the COPD assessment test (CAT) and the Clinical COPD questionnaire (CCQ) as PROMs used. Finally, a variety of disease-specific PROMs is used in mental health care, including the Generalised Anxiety Disorder tool (GAD-7), the Manchester Short Assessment for Quality of Life (MANSA), the Depression Anxiety Stress Scale (DASS), the Patient Health Questionnaire (PHQ-9), the Symptom Checklist 90 (SCL90), and the Hospital Anxiety and Depression Scale (HAD).

Generic PROMs were sometimes used alongside disease-specific instruments, to give a more complete picture of the outcomes of care. In these cases, the most common used generic PROMs instrument reported by national experts was the EuroQol (EQ-5D). Other PROMs used include Short Form (SF-36 and SF-12), the Health Assessment Questionnaire (HAQ), and PROMIS generic instruments. The EQ-5D was most commonly used in orthopedic surgery.

National experts emphasised that use of PROMs in chronic disease is more challenging than for elective surgery, with higher administrative costs and methodological challenges. Measuring health outcomes of patients with respect to care co-ordination and integrated care is rare. Some countries use PREMs for evaluating care co-ordination. Some examples of using PROMs in an integrated care approach in multidisciplinary settings.

In countries that are trialling PROMs, they are more commonly used to measure the outcomes of patients following an elective surgical intervention, particularly joint replacement procedures. There is much less information about the outcomes of patients with chronic disease, and patients in mental health care and long-term care. This is in part due to the fact that the ongoing nature of these conditions presents challenges in survey design and implementation. Given that these patients will pose a more significant burden on the health system in coming years, gaining an understanding of their outcomes will be critical. Additionally, very little is known about outcomes in the areas of emergency care, informal care, palliative care and preventive care.

Box 1.6. PROMs study design

Pre- and post-intervention studies have the benefit of measuring the effectiveness of a particular intervention. These studies have the advantage of temporality, although they do not take into account other factors that may influence changes in a patient's health outcomes. Therefore caution should be applied in attributing change solely to the intervention. There also remain unresolved questions as to at what point after an intervention the second measurement should be taken. While these studies are useful for specified interventions, their applicability to chronic disease is more limited. Examples of pre- and post-intervention studies are those used in hip and knee surgery patients.

Cross-sectional studies are limited in that generic questions on health status do not reveal much on their own. However, they provide a snapshot of comparisons between different groups at a point in time, and can form the basis of something to build upon. It is not possible to establish temporality because the survey captures a moment in time. For example, if a patient is experiencing anxiety and chronic pain, it is difficult to determine with a cross-sectional survey whether the chronic pain has caused the patient anxiety, or whether the anxiety has manifested in physical symptoms such as pain. An example of a cross-sectional study is the Commonwealth Fund international policy surveys measuring patient experience.

Longitudinal studies are a scientifically robust method of collecting information over time. They are suitable to measure the outcomes of chronic disease patients but, as discussed earlier, the design and implementation will be challenging when it comes to patients with multiple co-morbidities receiving care from several providers. Longitudinal studies do not have the temporality problems that cross-sectional studies have, as changes in a patient's condition can be detected over time. Questions remain around the frequency in which measurements should be taken. Longitudinal studies are more resource-intensive, and come with the risk of panel attrition due to participant drop-out or death. This raises the possibility of missing data. Examples of these studies include longitudinal PROMs studies assessing the outcomes of cancer patients over a period of time.

All three study designs come with the risk of selection bias, as the samples may not be representative of populations. There is also potential for information bias in the form of missing data, if patients do not complete the entire survey. There is the possibility of recall bias, if participants are asked to retrospectively recount symptoms they experienced. Variation in response rates across countries can also raise the risk of bias.

PROMs have the potential to inform decisions by policy makers with regards to resource allocation. New Zealand has adopted a "points system" to prioritise patients for cataract surgery, using a questionnaire measuring the impact on daily living. This PROMs information is used alongside the clinician's assessment about the improvement in health possible through treatment, as the basis for prioritising patients for cataract surgery (Cumming, 2015; Derret et al., 2013; Devlin et al., 2010).

In New Zealand national registries for joint replacement with PROMs being tracked longer-term, providers receive individual annual reports showing PROs for their patients compared against national averages. The Oxford Hip Score (OHS) and Oxford Knee Score (OKS), as well as adapted versions of the Oxford-12 score for other joints. If providers are concerned about their own performance they can request more detailed PRO information about their individual patients PROMs from the registry (NZOA, 2016).

In the United States, the Food and Drug Administration considers PROMs when assessing claims made by pharmaceutical companies. The Centers for Medicare & Medicaid Services (CMS) in the United States is moving towards a mandated outcomes-based payment model after hip and knee joint replacement surgery, including the use of PROMs. The programme started in April 2016 with a pilot project. Professionals will need to include PROMs in their clinical care and submit their results to meet new standards for reimbursement without incurring penalties. Although hospitals are not required to submit PROMs data; participation will result in higher reimbursements. PROMs included in the programme the HOOS, KOOS, VR12 and variants of these measures (CMS, 2016). Development of methodology for data collection and risk adjustment is included in the project (CORE, 2015). The American Joint Replacement Registry provides the infrastructure for the data collection (AJJR, 2016).

In Australia there is a now a renewed interest in the use of patient reported outcome measures (PROMs) and patient reported experience measures (PREMs) combined with a focus on integrated models of care (Chen, 2015). The Australian Health Outcomes Collaboration (AHOC) recently published policy report about the use of PROMs from an Australian perspective. The report concludes that it is to be hoped that the re-emerging health outcomes focus in Australia may act as a catalyst to integrate PROMs use in the various efforts that are already being made to improve the quality of health systems and hospitals (Sansoni, 2016).

The Canadian Institute for Health Information has produced a background document to facilitate the implementation of PROMs (CIHI, 2015). The report identifies several regional-level initiatives in Canada, but a standardised programme for routine PROMs collection and reporting does not exist. Four generic PROMs are considered for common use in PROMs initiatives across Canada: the SF family of instruments (such as the VR-12), the EQ-5D, the Health Utilities Index (HUI), and PROMIS Global Health. The VR-12 and EQ-5D were

identified as the most suitable generic tools for routine PROMs data collection and reporting. The report emphasises the need for a co-ordinated and standardised approach across Canada to support local, regional, national, and international comparisons.

In Denmark the coherence between using PRO-data in both direct treatment of patients as well as for quality development is a main focus. In 2016 a white paper was presented by the Danish Knowledge Center for User Involvement in Health Care (ViBIS) and TrygFonden examining how patient-reported health data can be used in the clinical work, as well as for quality improvement in the Danish health care system. A group of experts was established to evaluate which requirements, potentials and barriers that need to be addressed in order to implement PROs and PROMs systematically in the Danish health care system and to feed into the existing quality improvement efforts. The approach is to center the treatment around the patient, so that the questionnaires follow the course of treatment across sectors. Thus, the questionnaires must be nationally standardised. The National Danish PRO Secretariat and National PRO Working Group have been established based on the yearly economic deal between the national government, the regions and the municipalities. The secretariat is responsible for establishing a number of standardised and evaluated PRO questionnaires to be used nationwide for all patients. The first areas of concern are hip/knee osteoarthritis, apoplexy and anxiety/depression. Already there are projects aiming to incorporate PRO-data in out-patient care in the areas of epilepsy, prostate cancer and chemo therapy for breast cancer. Furthermore, PRO-data is used in general practice as a means of involving patients and targeting treatment, e.g. in treatment of depression and for blood pressure measurements.

Health system performance assessment has become an area of increasing interest within the European Union. In April 2014 the European Commission adopted a Communication to propose an EU agenda on effective, accessible and resilient health systems. As a result a European consortium has developed a consensus document in 2016 to facilitate the use of PROMs by EU Member States to measure heath system performance. The report highlights the importance of international quality comparisons as a means to identify best practice across countries, and to trigger quality improvement initiatives at a national level (FIPRA, 2016).

International comparisons of PROMs are almost entirely absent

While PROMs trials are occurring in OECD countries, little has been done at a cross-country level. ICHOM is a notable example of an organisation working on defining global standard sets of outcome measures across several medical conditions. To date it has completed sets for 21 conditions: coronary artery disease, localised prostate cancer, advanced prostate cancer, low back pain, cataracts, Parkinson's disease, depression and anxiety, cleft lip and palate, lung cancer, hip and knee osteoarthritis, stroke, pregnancy and childbirth, inflammatory bowel disease, overactive bladder, colorectal cancer, breast cancer, craniofacial microsomia, older persons, heart failure, dementia and macular degeneration (ICHOM, 2016). In May of 2016 ICHOM announced the Global Outcomes Benchmarking (GLOBE) Programme. Since then they have launched two pilot benchmarking programmes based on the Hip and Knee Osteoarthritis (HKO) and Cataracts (CAT) standard sets. As of October 2016, 25 hospitals were participating in the HKO pilot and 55 in CAT, with representation from ten countries. The aim of the pilots is to assess the feasibility of collecting outcomes data in different regulatory and technical environments. Following the pilots, ICHOM anticipates scaling up its benchmarking efforts to advance comparison between organisations and across countries. In this regard, there is an opportunity for the OECD to work in partnership with the organisation.

In the case of both PROMs and PREMs, a limitation of international comparisons is the variation in response rates across providers and countries. In the case of PROMs, the evidence to date suggests response rates are likely to be higher in pre- and post-intervention studies of elective surgical procedures, than in cohort studies of chronic disease in primary care. There is also a risk that the most vulnerable or severely-affected people are less likely to participate due to cognitive difficulties, poor health literacy, language barriers, or inability to access or use technology (e.g. smart phones or tablet computers). Additionally, exclusion criteria will have to apply. In a cohort study of long-term conditions in primary care, the exclusion rate of patients with epilepsy was 46.7%, and this was related to a high proportion of patients with learning difficulties (Peters et al., 2013a). Consideration needs to be given as to how to maximise participation of these groups. Annex C discusses technical issues around response rates in more detail.

Box 1.7. Measures of pain

Hawker et al. (2011) have provided an overview of pain measures in adult rheumatology populations. Their findings for the VAS and NRS are summarised below.

Visual analog scale for pain

The pain VAS is a continuous scale comprised of a horizontal (HVAS) or vertical (VVAS) line, usually 10 centimeters (100 mm) in length, anchored by two verbal descriptors, one for each symptom extreme. Instructions, time period for reporting, and verbal descriptor anchors have varied widely in the literature depending on intended use of the scale.

For pain intensity, the scale is most commonly anchored by "no pain" (score of 0) and "pain as bad as it could be" or "worst imaginable pain" (score of 100 [100-mm scale]). To avoid clustering of scores around a preferred numeric value, numbers or verbal descriptors at intermediate points are not recommended. Recall period for items varies, but most commonly respondents are asked to report "current" pain intensity or pain intensity "in the last 24 hours". The respondent is asked to place a line perpendicular to the VAS line at the point that represents their pain intensity. A higher score indicates greater pain intensity. The VAS is administered as a paper and pencil measure. As a result, it cannot be administered verbally or by phone. The pain VAS is available in the public domain at no cost.

Test–retest reliability has been shown to be good, but higher among literate than illiterate patients. In patients with chronic inflammatory or degenerative joint pain, the pain VAS has demonstrated sensitivity to changes in pain assessed hourly for a maximum of four hours and weekly for up to four weeks following analgesic therapy. In patients with rheumatoid arthritis, the minimal clinically significant change has been estimated as 1.1 points on an 11-point scale (or 11 points on a 100-point scale)

Numeric rating scale for pain

The pain NRS is a single 11-point numeric scale in which a respondent selects a whole number (0-10 integers) that best reflects the intensity of their pain. The common format is a horizontal bar or lin. Similar to the pain VAS, the NRS is anchored by terms describing pain severity extremes with 0 representing one pain extreme (e.g., "no pain") and 10 representing the other pain extreme (e.g., "pain as bad as you can imagine" and "worst pain imaginable"). Recall period for items varies, but most commonly respondents are asked to report pain intensity "in the last 24 hours" or average pain intensity.

The NRS can be administered verbally (therefore also by telephone) or graphically for self-completion. The respondent is asked to indicate the numeric value on the segmented scale that best describes their pain intensity. Higher scores indicate greater pain intensity. The pain NRS is available in the public domain at no cost.

High test–retest reliability has been observed in both literate and illiterate patients. For construct validity, the NRS was shown to be highly correlated to the VAS in patients with rheumatic and other chronic pain conditions. Analyses of the relationships between changes in pain NRS scores and patient reports of overall improvement demonstrated a reduction of 2 points, or 30%, on the pain NRS scores to be clinically important.

Source: Hawker, G.A. et al. (2011), "Measures of Adult Pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP)", *Arthritis Care & Research* (Hoboken), Vol. 63, Suppl. 11, pp. S240-252.

2. Wider benchmarking of patient-reported indicators of health system performance

It is apparent from the previous section that the international benchmarking of patient-reported indicators are in different stages of development. For PREMs, the issue is more one of reviewing progress that has been made so far, extending country coverage and moving to new sectors. For PROMs, work will be needed to get an international consensus on the information to be collected. Once the most promising sets of patient-reported indicators that have the potential to better drive international comparison of health system performance are identified, the key next steps will be to collect, analyse and publish them. This section provides recommendations from the HLRG on how the OECD could help lead international work on the regular, systematic benchmarking of a wider set of patient-reported indicators.

2.1. Criteria to determine scope of conditions and sectors for patient-reported indicators

In determining how the OECD should expand its PREMs programme and commence a PROMs programme, a number of criteria should be considered. Key among them is *actionability*. The collection of patient-reported metrics should assist in service evaluation, drive quality improvement, and inform decision making with regards to resource allocation. It should assist with identifying high variation in clinical practice and the potential to reduce waste.

The indicators should have strong *relevance* for health systems, and be meaningful for patients. Conditions with a high public health burden should therefore be prioritised. Of particular importance is the development of patient-reported metrics in chronic disease, assessing how well patients are engaging in self-management, and how well health systems are delivering integrated care particularly for people with multiple morbidities. They should capture the things that matter to patients, such as quality of life.

The *cost of implementation* is another important consideration for governments. Instruments that have been standardised and validated are already available in the public domain at no cost. The CAHPS Clinician and Group survey tools, for example, have been validated to measure patient experience in primary care. They are being used in several organisations, while others are incorporating core questions from the survey into their existing tools to move towards standardisation. Traditional mail and telephone surveys pose a higher cost burden, and the shift to web-based tools has potential to reduce the cost and improve the ease and speed of acquiring and using survey data (Browne et al., 2010). Electronic instruments come with the benefits of being interactive and minimising data entry errors, but come with high set-up costs.

The *availability* of measures and *feasibility* of data collection are also important considerations. It would be more feasible to begin with instruments that have already been validated, rather than introduce new instruments. The burden imposed upon patients should be low, and ease of use should be enhanced to maximise participation. Additionally, the amount of time providers spend on administering surveys should be minimised, to ensure their clinical time with patients is maximised. Measuring interventions that take place in the acute-care setting is generally more straightforward. However, in clinical areas where less is known, particularly chronic disease, the feasibility is more challenging.

In addition to the above criteria, instruments should be sensitive to detecting change in health status, and be assessed on their *psychometric properties*. This is discussed further in Section 2.3. The International Society for Quality of Life Research (ISOQOL) recommendations for minimum standards for PROMs is provided in Annex D.

Extending the international benchmarking of PREMs

In extending its work on PREMs, the HLRG recommended that the OECD should start by determining which conditions and sectors to focus on, based on the above criteria. Through the HCQI project, the OECD Health at a Glance reports four indicators of general population patient experience in ambulatory care. The Commonwealth Fund conducts surveys of both general population and more specific patient groups (Box 1.2). A key question for consideration is whether the OECD should focus on collecting PREMs relating to the general population or more specific patient groups and conditions, and in which sectors. In this regard, the HLRG recommended that the OECD start by building on existing survey instruments, and rapidly extend to conditions that to date

have received little attention. This would create new knowledge about the experience and outcomes of care in conditions where less is known.

The HLRG also advised that the OECD Secretariat should begin by expanding its PREMs work in ambulatory care. This would build upon the work of the Commonwealth Fund, which conducts population-based international surveys focusing on PREMs in 11 countries. The Commonwealth Fund surveys enable cross-country comparisons on issues that matter universally in health systems, and thus facilitate mutual learning between countries. There is potential for the OECD to build on this work by adding more indicators and more countries. The HLRG advised that the OECD should gather consensus to collect and report a core set of questions on patient experience with regards to care co-ordination and patient safety. The subsequent step should be to extend PREMs into acute care, with questions including patient safety and other measures. The PREMs programme should later extend to mental health care, long-term care, palliative care, emergency care, informal care and preventive care.

Extending the international benchmarking of PROMs

The HLRG addressed three issues in considering how to develop an internationally comparable set of data based on PROMs. First, which diseases and sectors should be covered, based on what instruments already exist and what instruments will need to be developed? Second, how frequently should the data be collected – before and after an intervention, as a cross-sectional survey, or on a longitudinal basis over time? Third, should the data be collected on a national basis for international comparison, or would an international survey developed for comparisons of service providers (e.g. hospitals) across different countries be of use?

On the first of these questions, the HLRG recommended that it would be most feasible for the OECD to begin by focusing on those diseases and interventions where trials are more advanced. There is also an opportunity for the OECD to eventually – once sufficient data are available – provide a time series to compare the speed of, and extent to which, patient outcomes are improving or deteriorating over time intervals across countries. This could facilitate mutual learnings by identifying countries that have achieved striking improvements.

The HLRG also recommended that it would be advantageous to begin by focusing on clinical areas were OECD already collects other

data, such as prevalence of risk factors, indicators of need, activity volumes or survival estimates. Such data could be used to interpret patient-reported indicators and place them in context.

With these two principles in mind, the HLRG advised that the OECD should gather consensus among OECD countries to implement standardised, validated instruments for hip and knee surgery. This would draw upon the work of groups that have already identified valid instruments, such as the International Consortium for Health Outcomes Measurement (ICHOM). Using established structures such as the Health Care Quality Indicators Expert Group, the OECD should explore the best options, taking into account what is most feasible and useful. Options include using the generic EQ-5D as well as the condition-specific Oxford Hip Score and Oxford Knee Score, as is the current practice in the NHS.

PROMs data collection is well established in hip and knee arthroplasty. A member survey of the International Society of Arthroplasty Registries (ISAR) registries showed that eight registries administered a PROMs programme that covered all elective hip or knee arthroplasty patients and six registries collected PROMs for sample populations (Rolfson et al., 2016). The most common generic instruments used were the EQ-5D, SF-12 or the VR-12. The most common specific PROMs were the HOOS, KOOS, OHS, OKS, WOMAC, and UCLA Activity Score. ISAR has also recommended best practices in the selection, administration, and interpretation of PROMs for hip and knee arthroplasty registries. Although the NHS PROMs initiative is not a formal clinical registry, PROMs data are collected at national level and can be added to the list of PROMs programmes that cover all elective hip or knee arthroplasty patients.

While it would be desirable for all countries to use the same instrument, this is improbable. An alternative could be for countries to use different instruments of equal rigour, and to map between these instruments. Mapping can be considered the development and use of a model or algorithm to predict health-related utility values. The key metric of interest is the quality-adjusted life year (QALY) for the purpose of economic evaluation of an intervention, to inform decisions about resource allocation. For example, it could be possible to map between two of the most commonly used generic instruments, the SF-36 and EQ-5D, as well as the condition-specific Western Ontario and McMaster Universities Arthritis Index (WOMAC) and the Oxford Hip Score and Oxford Knee Score.

The HLRG advised that the OECD should build on this work with the collection of PROMs data in patients with chronic disease. The ongoing nature of long-term conditions, coupled with additional psychosocial issues, make the collection of PROMs less straightforward. Yet with chronic disease, mental health conditions and dementia becoming more significant issues for health systems and consuming more health resources, learning more about the outcomes of these patients is essential. Any PROMs instruments that are used must be well-established and properly developed and validated, and reflect what matters to patients. Therefore, they should have undergone a process including focus groups to understand what measures of quality of life matter most to patients, bearing in mind that patients should be the principal beneficiaries of PROMs data collection.

Cancer treatment has a significant impact on patients' quality of life and PROMs are commonly used as outcome measures in cancer research (Efficace, 2014). Several countries have implemented registries for different types of cancer, and evidence shows that it is feasible to integrate PROs into routine cancer care, and that they improve process and outcomes of care. International collaboration has resulted in a robust "family" of PROMs developed via the European Organization for Research and Treatment of Cancer (EORTC). The EORTC Quality of Life Core Questionnaire, the QLQ-C30, is one of the most widely used cancer specific Health Related Quality of Life questionnaires in the world. It has been translated and linguistically validated into more than 90 languages and extended with over 40 validated modules for specific cancers (Fayers et al., 2002).

Standardised, validated PROMs instruments already exist for cancer, but there is little in the way of cross-country comparison. For example, prostate cancer is among the most common cancers in men, making this a rich and valuable area for PROMs instruments to be implemented across OECD countries.

Following cancer, the HLRG advised that the OECD should seek consensus to implement standardised, validated instruments across countries for other chronic conditions. Instruments should also be implemented for mental health care, long-term care, emergency care, palliative care, informal care and preventive care.

The HLRG also recommended that the focus should also shift to PROMs requiring a longer timeframe, for conditions where instruments are not well developed. Of particular importance are patients with

multiple co-morbidities. In these cases, it may be necessary for patients to complete more than one disease-specific survey. This imposes a high burden on patients who may not understand the value of completing such surveys, and raises the possibility of the need to develop PROMs instruments targeted to patients with multiple conditions.

The second question, on the frequency in which measurements should be taken in patients with long-term conditions, is a matter to which the answer depends on the disease or intervention which is being covered. For example, for joint replacements, it is clear that the question needs to be posed pre- and post-intervention. However, whether patients with chronic conditions should be asked to complete a survey each year, every two years, or less frequently, or whether they should be asked retrospective questions, is a matter that will need further investigation.

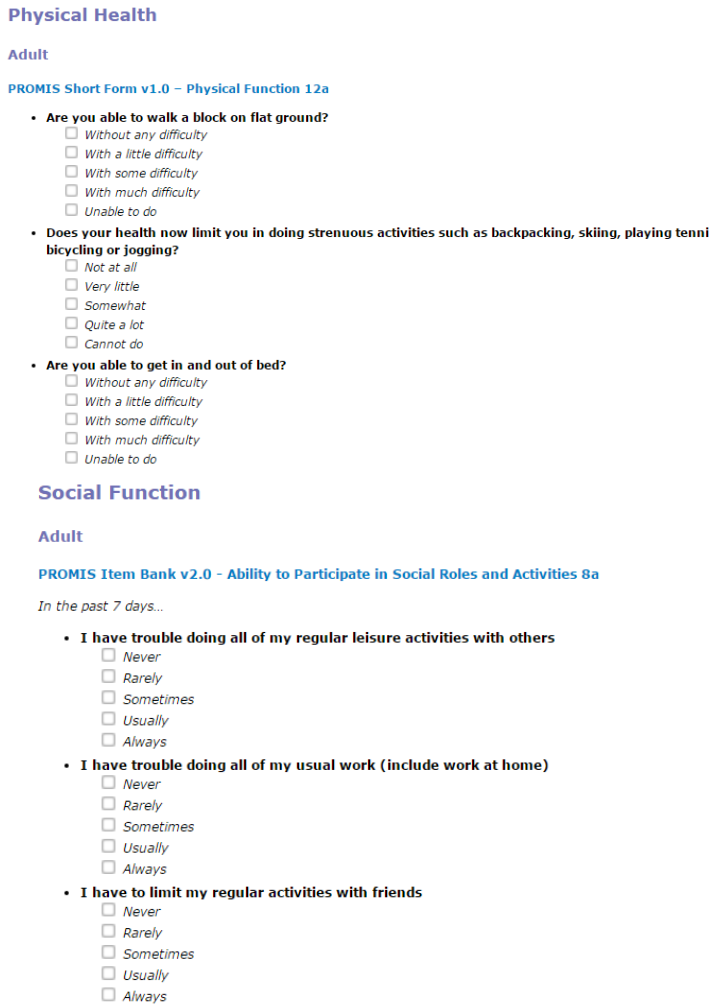
The third question is whether the OECD should seek to collect patient-reported indicators data at a whole-system level, which would facilitate cross-country comparisons, or whether it should also explore the possibility of collecting and reporting anonymised hospital-level data for a limited number of conditions or interventions, potentially in one or more sectors. This might require case-mix adjustment to make the comparisons across providers meaningful. The OECD is currently exploring the capacity for data collection, analysis and reporting on international variations in hospital-level performance as part of the Health Care Quality Indicators project. Countries that already report at a provider level include the NHS in England, which requires all hospitals to collect PROMs for four surgical procedures to enable hospital-level comparisons. In Sweden, hip arthroplasty patients report on pain and health-related quality of life, and this information is publicly reported at a regional and hospital level. Whether the OECD collects and reports anonymised hospital-level data will require further consideration.

Internationally comparable PROMs measures will be of most use if they use a combination of generic and disease-specific questions. Generic questions enable comparisons across conditions, which can assist in decisions about resource allocation. However, they lack the precision of disease-specific questions, which capture more information about conditions. In the case of generic questions, the EQ-5D measures health using different levels of severity to describe mobility, self-care, usual activities, pain and discomfort, and anxiety and depression.

Another option is the Patient-reported Outcomes Measurement Information System (PROMIS), funded by the National Institutes of

Health, which compiles a core set of questions and uses generic tests to assess common outcomes for a range of chronic diseases. PROMIS began in 2004 and aims to provide clinicians and researchers access to precise, valid, and responsive measures of health. PROMIS items measure pain, fatigue, emotional distress, physical functioning, social role participation, and general health for both adults and children (http://www.nihpromis.org/default#6) (Figure 2.1).

Figure 2.1. PROMIS generic sample questions



Source: PROMIS, *http://www.nihpromis.org/Measures/SampleQuestions* (accessed 27/08/2015).

For disease-specific PROMs, the HLRG recommended that the OECD consider exploring opportunities to build on the work of the International Consortium for Health Outcomes Measurement (ICHOM), which has developed standard sets for several conditions. For example, in the case of localised prostate cancer, measures of patient-reported health status include vitality, sexual dysfunction, bowel irritation, urinary frequency obstruction irritation, and urinary incontinence (ICHOM, 2015b).

2.2. Standardisation and validation of patient-reported indicators across countries

To ensure cross-country comparability of data, there is a need to use standardised, validated instruments. Instruments should be relevant to patients, and acceptable to both physicians and patients. For a standardised approach to patient-reported indicators, the HLRG advised that the OECD should ensure that their selection is based on strict criteria, which includes taking patients' priorities into account using focus groups. Instruments should meet methodological requirements of cognitive testing and psychometric properties of validity, reliability and international comparability.

Psychometric analysis is commonly undertaken, although the extent to which OECD countries do this varies. In considering the reliability and validity of instruments, the NHS experience with cataract surgery provides a practical example. Cataract surgery is one of the most common procedures performed in hospitals, and the ageing population will make this a more substantial public health issue in coming years. There would therefore be great benefit to learning more about the outcomes of these patients.

Cataract surgery was to have been included in the NHS PROMs programme, but was abandoned due to concerns about methodology (Browne et al., 2007). This came after a study examining the outcomes of patients following cataract surgery using the VF-14 test, which measures visual function, raised questions about the validity of the instrument, in part due to "response shift" (a change in an individual's values, standards and perception of quality of life). Patients may be unaware of any visual dysfunction before surgery, because any deterioration could be gradual. In some cases, patients may become aware of how much clearer the world could be only after surgery. Such patients report no or little dysfunction before surgery and the same

afterwards, but still report that the operation has been beneficial (Black et al., 2009). Taking a measurement before and after an intervention is not necessarily useful in these cases. Other PROMs for cataract surgery are being used. For example, in its Standard Set for Cataracts, ICHOM recommends the Catquest 9 SF. Annex D provides the International Society for Quality of Life Research (ISOQOL) recommendations for minimum standards for PROMs.

A standardised approach should also be adopted with regards to data collection, to ensure comparability. The HLRG considered whether the OECD should collect aggregated data from countries, or whether it should engage in primary data collection, as it does for the educational outcomes survey PISA. It is likely to be more feasible for the OECD to collect a representative sample of patients, as does the Commonwealth Fund. Therefore, in scoping patient-reported indicators work, consideration should be given as to how to identify appropriate samples.

In countries where primary data is available via registries such as England, Sweden, the United States and the Netherlands, it would be feasible to use the full database. Methodological issues in using different types of samples should be considered.

Knowing that current data collection methods vary considerably between and within countries it is important that OECD develops instructions for standardisation of data collection to ensure minimum data quality, but also providing sufficient latitude for countries to allow for tailoring to national circumstances.

In developing its national PROMs programme, England has standardised procedures for data collection at the national level to ensure data consistency, permitting health services to be meaningfully benchmarked. Its data collection methodology draws on research it commissioned from the London School of Hygiene and Tropical Medicine, which piloted a small number of procedures with 2 400 patients at 24 sites.

To promote a consistent approach, the Department of Health sets out the respective responsibilities of all parties in the collection of PROMs. Providers administer questionnaires to patients, and must ensure the collected data are as representative of their patient populations as possible. They are expected to use minority language versions of PROMs questionnaires where necessary. Commissioners work with providers to establish appropriate thresholds for the participation rate and to hold them to account where performance does not meet the agreed

levels. The PROMs Administration contractor collects the data and converts the data into an electronic record for transmission to the NHS Information Centre (IC), which links the identifiable record-level PROMs data to existing routinely collected administrative data. To produce comparable aggregated data, the PROMs data aggregation contractor develops evidence-based case-mix and risk-adjustment methodology, which is applied to the linked, record-level data.

A further option for consideration is the feasibility of a combined patient-reported indicators survey, to minimise the burden for patients and providers. However, designing a survey relevant to patients with a range of conditions is challenging. Another challenge is taking into account that PREMs surveys tend to be cross-sectional, whereas PROMs surveys are usually pre- and post-intervention studies for surgical procedures, or longitudinal studies for chronic disease. An attempt at a combined patient-reported indicators survey was made in the NHS in England, with the Outcomes and Experiences Questionnaire (Annex E). The survey has not yet been adopted in the NHS.

To facilitate international comparisons of health outcomes, crosswalk algorithms between scores on different PROMs may be a potential avenue. Several efforts have explored how scores collected using one questionnaire may be converted into comparable scores for a different related questionnaire via crosswalks (Brazier et al., 2010; Wu et al., 2005; Chan et al., 2012; Le, 2014; Choi et al., 2012; Oude Voshaar et al., 2014; Bujkiewicz et al., 2014). Many of these studies were aimed at converting scores from disease specific PROMs to generic PROMs such as the EQ-5D of DF-6D. The rationale for these crosswalks is that most disease specific PROMs cannot be used in cost-effectiveness analysis using cost per quality adjusted life year (QALY) (Brazier et al., 2010). We did not identify published examples of crosswalks between any of the disease specific instruments in hip osteoarthritis.

A crosswalk algorithm is available to convert SF-12 responses to EQ-5D index scores, which may enable comparisons between the tools (Le, 2014). The authors used a probabilistic mapping approach on EQ-5D utility scores based on SF-12 responses using Bayesian networks (Le, 2014). Crosswalk algorithms have also been established between the mental components of the PROMIS global health and the VR-12 (Cella et al., 2012).

Future research and development around comparing scores from different PROMs would explore the standardisation of scores by

transforming specific PROMs scores to a generic metric, such as the T-score metric developed by the Patient-reported Outcomes Information System (PROMIS) (Cella et al., 2010). A T-score is a standardised score, like z-scores and IQ scores. All standardised scores have a "middle" score; it is zero for z-scores, 100 for IQ scores, and 50 for T-scores. In a T-score metric 50 is the mean of a relevant reference population and 10 is the standard deviation (SD) of that population.

In Dutch mental health care, changes in T-scores are used for benchmarking of outcomes of mental health care services. T-scores are estimated based on raw scores of PROMs instruments, resulting in a mean of 50 and a standard deviation of 10. In the next step, T-scores are transformed to normalised T-scores, by regressing raw T-scores onto percentile ranks. Normalised scores have a true interval scale and a normal distribution which makes subtraction of pre-and-post T-scores permissible. The change in T-score is calculated as the pre-test T-score minus the post-test T-score for each patient. The difference score is the prime indicator of the performance of mental health care institutes (de Beurs, 2016).

Further research would assess the validity and reliability of using crosswalk algorithms or standardisation of scores such for enhancing the interpretability and comparability of outcomes from generic and disease specific PROMs (Johnston et al., 2013). Measurement theories such as the Item Response Theory (IRT) and Differential Item Functioning (DIF) may be useful in developing methodology for comparing and combining scores measured with different PROMs across countries and cultures (Chan et al., 2012; McHorney et al., 2006; Choi et al., 2012).

2.3. Engaging stakeholders in the implementation of patient-reported indicators

In order to maximise the potential of patient-reported indicators, patients, their carers and clinicians need to be educated about the benefits. The implementation of such programmes will be successful only if those doing the work see the value. Therefore, stakeholder engagement is a critical component of the implementation of any PREMs or PROMs programme.

In the survey and set of structured interviews commissioned by the HLRG, national experts emphasised that engaging clinicians is crucial to the success of a PROMs programme. The closer the PROMs work is to the clinician's practice the better, as they will see the value if they can

use it to improve patient care. On this basis, clinicians are less likely to favor using generic PROMs. If clinicians are not measuring outcomes that help their patient's treatment plan, then they are unlikely to use them. Also for patients PROMs instruments need to be relevant to maximise participation. Patients will not bother answering questions that are not relevant to them. On this basis, patients are more likely to favor disease-specific tools as generic tools may be less relevant.

Interviewees also noted that stimulating PROMs measurements can be achieved via bottom-up and top-down approaches. Most countries have started with a voluntary bottom-up approach, and mandatory participation is rarer. In England, for example, the use of PROMs by providers was mandated in 2009 for several elective surgical procedures. Motivation of clinician and patients for using PROMs in clinical practice (bottom-up) is important, although examples from the Netherlands and Sweden show that economic incentives (top-down) can encourage provider participation.

Finally, interviewees also stressed the importance of distinguishing the different purposes of PROMs, to anticipate on potential differences between OECD countries in the collection of PROMs data across its member countries. Research has shown that the use of PROMs in clinical practice and for performance measurement has developed separately and in parallel. Data collection approaches that support use of PROs in health care are underdeveloped, need better integration with clinical care, and will need to be tailored to the characteristics of the healthcare system.2 These approaches across may lead to a tailored approach in data collection across OECD countries, keeping in mind that data should be comparable for cross-country comparisons for which a standardised approach is required.

In the case of patients, seeking their views and involving them in the design of survey instruments is essential. This should begin with asking patients what matters to them beyond survival, to take in the measures of quality of life they consider fundamental. Establishing focus groups elevates patients to the role of partners, and gives them a say in what indicators are collected. Importantly, they can also provide insight into the most practical and less onerous way of collecting data, thus reducing the burden and maximising participation. Carers may also assist in this process, particularly if they are needed to provide proxy reporting. Fully engaging patients in the process encourages them to take ownership of their health care and play an active role in self-management. Closing the feedback loop and sharing information obtained from patients with them provides an incentive for their participation.

Engaging clinicians will also be a key part of the process. Imposing participation on clinicians may cause resentment, if they cannot see the value. Additionally, there is lingering scepticism among some clinicians as to the benefits of collecting the information. In a qualitative study of surgeons, for example, there were mixed views on the value of peer-benchmarked PROMs data. Some were reassured that their practice was similar to that of their peers. However, they considered PROMs information alone insufficient to help identify opportunities for quality improvements. Some expressed concern about the scientific properties of PROMs, and considered the data subjective and therefore less trustworthy. There was also confusion in the difference between measures of patient outcomes and patient experience and satisfaction. Surgeons were also concerned about the accuracy of PROMs, related to possible biases, confounding factors, and chance. They reported difficulty making sense of the PROMs feedback, and using it to identify opportunities for quality improvement. Additionally, there were concerns that data collection would add to their workload (Boyce et al., 2014).

This indicates a clear need to educate clinicians in how to translate the metrics reported by patients into meaningful changes in clinical practice that will enhance patient outcomes. For instance, studies have also demonstrated that the use of PROMs can improve communication between clinicians and patients. The key will be to train clinicians in the optimal use of that data. Involving clinicians in the process can also be useful in identifying ways to reduce the data collection burden and other barriers to participation.

Identifying and overcoming barriers

In the survey and set of structured interviews commissioned by the HLRG, national experts identified many challenges for national approaches including validity and reliability of data collection, the use of standardised instruments, and case-mix adjustment. In particular, case-mix or risk adjustment has been a problem for many countries. Many countries are struggling with case-mix adjustment e.g. due to lack of robustness of data. However, case-mix adjustment is considered important for provider engagement if data is to be used for public reporting. Nevertheless, national respondents showed enthusiasm for the OECD to build consensus in the selection of PROMs instruments for international data collection and comparisons.

In Canada, At CIHI's PROMs Forum held in February 2015, potential barriers to PROMs data collection identified included limited

resources for data collection, engagement and buy-in of clinicians and administrative stakeholders, the ability to collect data (e.g., pre- and post-intervention) and reaching an agreement on common tools. It has also been identified that understanding how to appropriately use and interpret PROMs is important (e.g., must be actionable and relevant). Furthermore, standards for administration (e.g., timing and frequency of collection) will need to be developed to ensure applicability in the specific clinical area/sector (e.g., surgery versus chronic disease).

In Israel, there were initially technology challenges as the aim from the start was to integrate PROMs with the electronic medical records and patient portals. Following technology development, challenges have been getting clinicians engaged and getting patients to report the follow-up data.

In the Netherlands, the main challenges have been in the use of PROMs are setting up a routine of measurement and in providing infrastructure for extracting data from electronic health records or separate software applications to (national) databases. The next challenge will be to derive valid and reliable quality indicators from the data as collected, using proper case-mix adjustment and handling of missing data.

According to the Canadian policy survey response, major barriers to international comparability include selection of common survey tools/questions in survey tools; standardised measures/indicators for comparisons and ensuring these measures/indicators are relevant and actionable; the lack of a central repository of data for international comparisons; sensitivities to local social/cultural differences; development of standard collection protocols (e.g., timing, frequency, sampling); varying privacy and legal requirements as well as differing levels of sensitivity to personal health information which may impact the ability and method for collecting patient reported measures.

According to England's policy survey response, barriers to comparability include variability of care pathways, and clinical definitions and likely collection modes (who asks whom at what point of their care) all of which can influence the responses given. Another factor is standardisation of the questions and response patterns across language and cultural bias. This means overall, there is difficulty in like-for-like comparison. Just as variation in outcomes within providers is larger than the variation across providers, the variation in outcomes within health systems is likely to be much bigger than across. This makes it difficult to interpret the data and draw meaningful conclusions.

Note

1. In January 2016 a survey concerning the use of patient-reported health system performance indicators was sent to OECD member states. Thirteen countries responded. In addition, eleven interviews were conducted with experts from the countries that participated to the survey.

3. Supporting countries to embed patient-reported indicators into national health system performance assessment

As mentioned in Section 1, a difficulty with patient-reported indicators has been the low response rate from patients. Patient participation can be maximised with surveys that are easy to use. OECD countries are using technology to facilitate more easily comprehensible and prompt survey response. There is value in exploring this further.

Several data sources (self-report vs, proxy/observer), modes (self-administration, interviewer-administration), methods (paper-pencil, computer, telephone) and settings (home, clinic) for PROMs data collection exist, which should be considered in data collection for cross-country comparisons.

The National Quality Forum (NQF) in the United States. has published detailed reports about methodological issues in PROMs data collection (Cella et al., 2012; Deutsch et al., 2012) and the NWS Agency for Clinical Innovation has also published a scoping review related to data collection issues (Chen, 2015). In addition, data sources for PROMs collection and storage may vary from local stand-alone databases and software, electronic health records, clinical registries, etc.

Wu et al. (2013) recently provided case-studies of a number of health care organisations in the United States that are now integrating PROMs data collections with the electronic health record to both promote quality improvement in clinical practice and for research concerning the effectiveness of interventions.

In the survey and set of structured interviews commissioned by the HLRG, national experts noted that the methods of PROMs data collection varied across the countries with paper-based data collection as most often used data collection method (n=10 countries), followed by mobile apps, tablets and/or computers (n=7). In six countries the data collection was embedded in electronic health records, and in

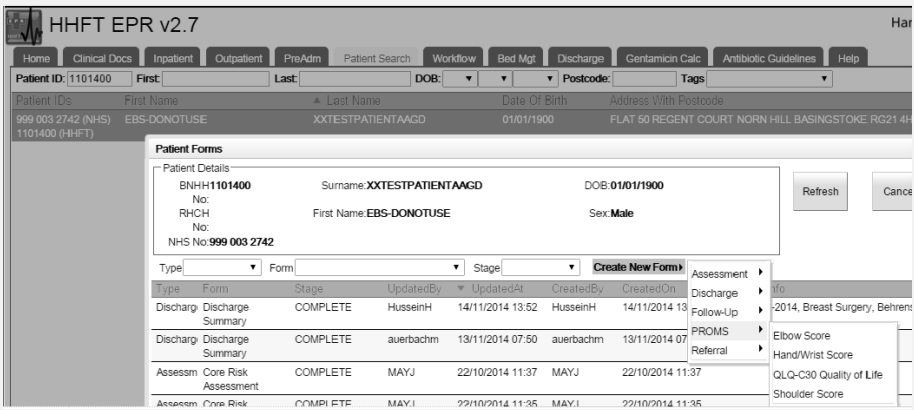
five countries data collection was embedded in clinical registries (see Table 3.1).

The HLRG stressed that there is potential for the OECD to assist countries in embedding patient-reported indicators into electronic health records, clinical registries, mobile apps and other sources. The OECD should start by examining the different methods already used in OECD countries. Co-operation would be sought from expert groups. Emphasis should be on how to translate the evidence of developed patient-reported indicators into the practice of embedding them in the data collection taking place within the information infrastructure of countries. The OECD should facilitate mutual learning between countries on these embedding processes (Box 3.1).

Box 3.1. Embedding patient-reported indicators into electronic health records, tablets and registries

While this work is in its infancy, some countries are exploring ways of embedding PROMs into electronic patient records. A recent presentation by the Hampshire Hospitals Foundation Trust (HHFT) in the NHS demonstrates how this work is being conducted (Figure 3.1). The HHFT previously did not have a structured approach to PROMs, with departments each doing things differently. Surveys were mainly paper based. As a result, the service formed a steering committee to develop a more consistent approach to PROMs across the service. The HHFT is trialling PROMs in the areas of orthopaedics, cancer services and gynaecology. It is being trialled in desktops, laptops and tablets (Figure 3.2) (Fokke and Simon, 2014).

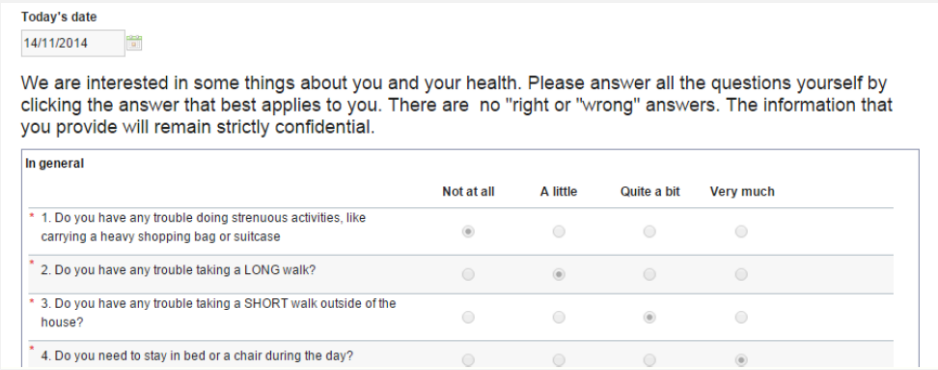
Figure 3.1. Embedding PROMs into electronic patient records at HHFT



Source: Fokke and Simon (2014).

Box 3.1. Embedding patient-reported indicators into electronic health records, tablets and registries *(cont.)*

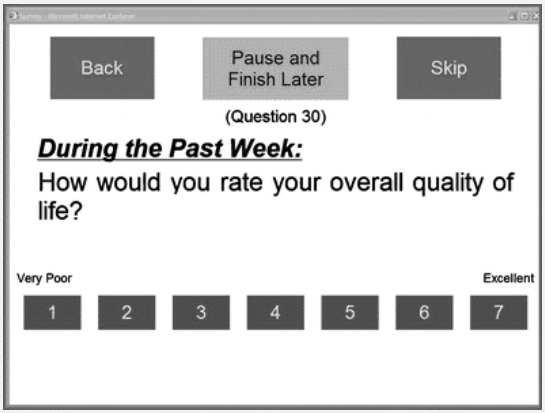
Figure 3.2. PROMs in cancer services: HHFT QLQ-C30



Source: Fokke and Simon (2014).

A US study demonstrates how PROMs can be used in tablets to maximise ease of use for elderly cancer patients (Figure 3.3). The survey comprises 30 questions from the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire C30 (QLQ-C30), designed to measure the quality of life of patients with cancer. It has a numeric rating scale of 0 (none) to 10 (worst possible) for the severity of a range of symptoms such as pain and nausea. It was designed to account for patients with poor vision (Fromme et al., 2011).

Figure 3.3. Example question page of the QLQ-C30 PROM used in a tablet



Source: Fromme et al. (2011).

Box 3.1. Embedding patient-reported indicators into electronic health records, tablets and registries *(cont.)*

In the United Kingdom, a mobile phone-based advanced symptom management system has been used to evaluate chemotherapy-related toxicity in patients with lung, breast or colorectal cancer. In a study, patients were asked to complete the electronic symptom questionnaire on their mobile phone, take their temperature using an electronic thermometer and enter this value into the mobile phone. The information was immediately sent in real time to the study server. A risk model was developed and incorporated into the study software, and was used to alert health professionals of any incoming symptom reports of concern. After completing the electronic symptom questionnaire, patients received self-care advice on their mobile phone related to the severity of the symptoms they reported. Patients reported improvements in communication with health professionals, the management of their symptoms, and feeling reassured their symptoms were being monitored while at home (McCann et al., 2009).

Clinical registries provide a potential source of PROM data collection. In Sweden, all public and private orthopaedic units that perform hip replacements participate on a voluntary basis in the Swedish Hip Arthroplasty Register. Individual data on diagnoses, laterality, and detailed information on implants and fixation are reported. A standardised protocol including PROMs was gradually introduced in Sweden in 2002. All patients are asked to complete a self-administered 10-item questionnaire pre-operatively and repeated at one, six, and ten years after surgery. Pre-operatively, patients complete the questionnaire either via a touch-screen application or paper questionnaire at the hospital clinic. At the three follow-up times, the questionnaire is mailed to patients. Non-respondents receive the first and only reminder after eight weeks. The response rate has been between 86% and 90% (Rolfson et al., 2011).

In implementing electronic patient-reported indicators, a key limitation is that few countries have implemented electronic health across all providers and sectors. Additionally, disease registries are sparse in some countries. Patients' fears regarding privacy also need to be allayed with strong safeguards and legislation. Legislation around data collection and privacy differs across OECD countries, which could present a barrier to a standardised approach. The OECD is already working with health and privacy authorities to develop a set of agreed principles around the use of sensitive personal health data. This work will be instrumental to efforts to improve measurement of patient outcomes.

Another consideration is whether electronic tools should be disease-specific or generic. While the former has the advantage of being developed for particular conditions, it cannot be applied across health services for other conditions. Some hospitals may wish to implement one tool across the service. There is also the possibility of bias being introduced in the instrument if electronically-administered surveys do not meet psychometric requirements (ICHOM, 2014). However, electronic surveys

have the benefit of being less time-consuming than paper surveys. Paper surveys also come with a risk of the introduction of errors when entering data, and missing values could compromise data quality. While electronic surveys can reduce these risks, the use of the internet alone may not provide a sufficient response rate in disease-specific populations, and could be supplemented with traditional paper questionnaires for people who do not have internet access (Rolfson et al., 2011).

In the case of registries, starting a new register requires a commitment by the profession to support the register and collect data, a strong core team to develop the register, and the inclusion of a limited number of data items (Rolfson et al., 2011). Given the emerging use of electronic PROMs, an ICHOM working group has established some minimum requirements for their use (Box 3.2).

Box 3.2. ICHOM Working Group minimum requirements for electronic PROM tools

The first level of requirements relates to the solution provider, the legal entity commercialising the tool:

1. Independence: The solution provider should be free from conflicts of interest, and clinicians should avoid solution providers that might be interested in using the raw data for commercial purposes.

2. Business continuity: The solution provider should be able to demonstrate the ability to sustain business operations for several years to ensure continuity of PROMs collection.

3. Maintenance: The solution provider should commit to providing maintenance for the tool to promptly deal with problems or evolutions.

4. Compliance with law: The solution provider should comply with all local and national laws.

There are additional requirements at the tool level:

1. Data ownership: The care provider should be the sole owner of data.

2. Data access: The care provider should have direct and unlimited access to the raw data. Any access to the data by the solution provider should be pre-approved by the care provider.

3. Data security: Data in transit between systems should be encrypted. Access to data must be recorded for audit purposes.

4. System reliability: A Service Level Agreement should define how the tool will be maintained and the response level to be expected. The tool should be available any time for patients, with little or no delay. Bugs and problems should be fixed quickly.

5. Unique patient identification: Since PROMs are typically tracked over time for the same patient, providers should be able to identify each patient and their records uniquely in the system.

6. System adaptability: The tool should be customisable by care providers. This can help to reduce bias and ensure replicability and comparability of results.

Source: ICHOM (2014).

References

ACSQHC – Australian Commission on Safety and Quality in Health Care (2014) "Framework for Australian Clinical Quality Registries", Australian Commission on Safety and Quality in Health Care, Sydney.

AJRR – American Joint Replacement Registry (2016), *AJJR's Patient-Reported Outcome Measures Guide*, American Joint Replacement Registry (AJRR), Rosemont, United States.

Aligning Forces for Quality (2010), *Good for Health, Good for Business: The Case for Measuring Patient Experience of Care*, Robert Wood Johnson Foundation, Washington, DC.

Alviar, M.J. et al. (2011), "Do Patient-reported Outcome Measures in Hip and Knee Arthroplasty Rehabilitation Have Robust Measurement Attributes? A Systematic Review", *Journal of Rehabilitation Medicine*, Vol. 43, No. 7, pp. 572-583.

Arden, N.K. et al. (2011), "What Is a Good Patient Reported Outcome After Total Hip Replacement?", *Osteoarthritis Cartilage*, Vol. 19, No. 2, pp. 155-162.

Argimona, J.M. et al (2004), "Health-related Quality of Life in Carers of Patients with Dementia", *Family Practice*, Vol. 21, No. 4, pp. 454-457, http://dx.doi.org/10.1093/fampra/cmh418.

Bakshi, A.B. et al. (2012), "Validation of the Care Transition Measure in Multi-ethnic South-East Asia in Singapore", *BMC Health Services Research*, Vol. 12, No. 256, http://dx.doi.org/10.1186/1472-6963-12-256.

Basser, M.R. (2015), "Benefits Case Study for Patient Reported Outcome Measures (PROMs) Outputs – Improving health outcomes for patients undergoing knee replacement, hip replacement, varicose vein and groin hernia treatments", Health & Social Care Information Centre, Version 1.3.

Bausewein, C. et al (2005), "Validation and Clinical Application of the German Version of the Palliative Care Outcome Scale", *Journal of Pain and Symptom Management*, Vol. 30, No. 1, pp. 51-62, http://dx.doi.org/10.1016/j.jpainsymman.2005.01.017.

Bellamy, N. (2002), "WOMAC: A 20-year Experiential Review of a Patient-centered Self-reported Health Status Questionnaire", *Journal of Rheumatology*, Vol. 29, No. 12, pp. 2473-2476.

Black, N. et al. (2009), "Is There Overutilisation of Cataract Surgery in England?", *British Journal of Ophthalmology*, Vol. 93, No. 1, pp. 13-17, http://dx.doi.Org/10.1136/bjo.2007.136150.

Bosworth, A. et al (2015), "Development and Validation of a Patient Reported Experience Measure (PREM) for Patients with Rheumatoid Arthritis (RA) and other Rheumatic Conditions", *Current Rheumatology Reviews*, Vol. 11, No. 1, pp. 1-7, http://dx.doi.org/10.2174/1573397111666150522093712.

Boyce, M.B. and J.P. Browne (2015), "The Effectiveness of Providing Peer Benchmarked Feedback to Hip Replacement Surgeons Based on Patient-reported Outcome Measures – Results from the PROFILE (Patient-Reported Outcomes: Feedback Interpretation and Learning Experiment) Trial: A Cluster Randomised Controlled Study", *BMJ Open*, Vol. 5, No. 7:e008325.

Boyce, M.B. and J.P. Browne (2013), "Does Providing Feedback on Patient-reported Outcomes to Healthcare Professionals Result in Better Outcomes for Patients? A Systematic Review", *Quality of Life Research*, Vol. 22, No. 9, pp. 2265-2278.

Boyce, M.B., J.P. Browne and J. Greenhalgh (2014), "Surgeon's Experiences of Receiving Peer Benchmarked Feedback Using Patient-reported Outcome Measures: A Qualitative Study", *Implementation Science*, Vol. 9, No. 84, http://dx.doi.org/10.1186/1748-5908-9-84.

Boyer, L. et al. (2013), "Evaluating the Impact of a Quality of Life Assessment with Feedback to Clinicians in Patients with Schizophrenia: Randomised Controlled Trial", *British Journal of Psychiatry*, Vol. 202, No. 6, pp. 447-453, http://dx.doi.org/10.1192/bjp.bp.112.123463.

Brazier, J.E. et al. (2010), "A Review of Studies Mapping (or Cross Walking) Non-preference Based Measures of Health to Generic Preference-based Measures", *European Journal of Health Economics – HEPAC: Health Economics in Prevention and Care*, Vol. 11, No. 2, pp. 215-225.

Browne, J. et al. (2007), "Patient Reported Outcome Measures (PROMs) in Elective Surgery: Report to the Department of Health", London School of Hygiene & Tropical Medicine.

Browne, K. et al. (2010), "Measuring Patient Experience as a Strategy for Improving Primary Care", *Health Affairs*, Vol. 29, No. 5, pp. 921-925, http://dx.doi.org/10.1377/hlthaff.2010.0238.

Bujkiewicz, S. et al. (2014), "Use of Bayesian Multivariate Meta-analysis to Estimate the HAQ for Mapping onto the EQ-5D Questionnaire in Rheumatoid Aarthritis", *Value Health*, Vol. 17, No. 1, pp. 109-115.

Carpenter, I. and J.P. Hirdes (2013), "Using InterRAI Assessment Systems to Measure and Maintain Quality of Long-term Care", *A Good Life in Old Age? Monitoring and Improving Quality in Long-term Care*, OECD/European Union, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264194564-7-en.

Cella, D. et al. (2012a), "Methodological Issues in the Selection, Administration and Use of Patient-reported Outcomes in Perfomance Measurement in Health Care Settings", National Quality Forum (NQF), Washington, DC.

Cella, D. et al. (2012b), "PROMIS Global Health – Mental Health Component and VR-12-Mental Component (Algorythmic Scores)", Northwestern University, Chicago.

Cella, D. et al. (2010), "The Patient-Reported Outcomes Measurement Information System (PROMIS) Developed and Tested its First Wave of Adult Self-reported Health Outcome Item Banks: 2005-2008", *Journal of Clinical Epidemiology*, Vol. 63, No. 11, pp. 1179-1194.

Chan, K.S. et al. (2012), "Measurement Equivalence in ADL and IADL Difficulty Across International Surveys of Aging: Findings from the HRS, SHARE, and ELSA", *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, Vol. 67, No. 1, pp. 121-132.

Chen, J. (2015), "Integrated Care: Patient Reported Outcome Measures and Patient Reported Experience Measures – A Rapid Scoping Rreview", NSW Agency for Clinical Innovation, Chatswood, Australia.

Chen, J., L. Ou and S.J. Hollis (2013), "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 Services Research*, Vol. 13, No. 211.

Choi, S.W. et al. (2012), "PROSetta Stone Analysis Report: A Rosetta Stone for Patient Reported Outcomes. Volume 1", Northwestern University, Chicago.

CIHI – Canadian Institute for Health Information (2015), "PROMs Background Document".

CMS – Centers for Medicare & Medicaid Services (2016) "Comprehensive Care for Joint Replacement Model", https://innovation.cms.gov/initiatives/CJR (accessed August 1, 2016).

Coleman, E.A., E. Mahoney and C. Parry (2005), "Assessing the Quality of Preparation for Posthospital Care from the Patient's Perspective: The Care Transitions Measure", *Medical Care*, Vol. 43, No. 3, pp. 246-255, http://dx.doi.org/10.1097/00005650-200503000-00007.

CORE – Center for Outrcomes Research and Evaluation (2015), "Patient-reported Outcomes Following Elective Primary Total Hip and/or Knee Arthroplasty: Hospital-level Performance Measures. Phase 3: Measure Methodology Report", CORE, New Haven, Yale.

Cumming, J. (2015), "Health Economics and Health Policy: Experiences from New Zealand", *Applied Health Economics and Health Policy*, Vol. 13, No. 3, pp. 281-289.

Davis, A.M. et al. (2009), "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*, Vol. 17, No. 7, pp. 843-847.

Davis, A.M. et al. (2008), "The Development of a Short Measure of Physical Function for Hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT Initiative", *Osteoarthritis Cartilage*, Vol. 16, No. 5, pp. 551-559.

de Beurs, E. et al. (2016), "Comparing Methods to Denote Treatment Outcome in Clinical Research and Benchmarking Mental Health Care", *Clinical Psychology & Psychotherapy*, Vol. 23, No. 4, pp. 308-318.

Department of Health (2008), "Guidance on the Routine Collection of Patient Reported Outcome Measures (PROMs) for the NHS in England 2009/10".

Derrett, S. et al. (2003), "Prioritizing Patients for Elective Surgery: A Prospective Study of Clinical Priority Assessment Criteria in New Zealand", *International Journal of Technology Assessment in Health Care*, Vol. 19, No. 1, pp. 91-105.

Deutsch, A. et al. (2012), "Patient-Reported Outcomes in Performance Measurement", National Quality Forum (NQF), Washington, DC.

Devlin, N.J. et al (2010), *Getting the Most out of PROMs: Putting Health Outcomes at the Heart of NHS Decision-Making*, The King's Fund.

DICA – Dutch Institute for Clinical Auditing (2016), "Jaarrapportage 2015", DICA, Leiden.

Efficace, F. et al. (2014), "Overcoming Barriers to the Implementation of Patient-reported Outcomes in Cancer Clinical Trials: The PROMOTION Registry", *Health and Quality of Life Outcomes*, Vol. 12, No. 86.

"EuroQol – A New Facility for the Measurement of Health-related Quality of Life" (1990), *Health Policy*, Vol. 16, No. 3, pp. 199-208.

Fayers, P. and A. Bottomley (2002), "Quality of Life Research Within the EORTC-the EORTC QLQ-C30. European Organisation for Research and Treatment of Cancer", *European Journal of Cancer*, Vol. Vol. 38, Suppl. 4, pp. S125-133.

FIPRA (2016), "Enhancing Value in European Health Systems: The Role of Outcomes Measurement", Brussels.

Fokke, C. and D. Simon (2014), "ePROMs Integration with the Electronic Patient Record, Hampshire Hospitals NHS Foundation Trust", Health Care Conferences UK, available at: http://www.healthcareconferencesuk.co.uk/news/newsfiles/chris-fokke-and-dominic-simon_861.pdf (accessed 14/08/2015).

Fremont, A.M. et al. (2001), "Patient-centered Processes of Care and Long-term Outcomes of Myocardial Infarction", *Journal of General Internal Medicine*, Vol. 16, No. 12, pp. 800-808, http://dx.doi.org/10.1111/j.1525-1497.2001.10102.x.

Friedberg, M.W. et al. (2011), "Physician Groups' Use of Data from Patient Experience Surveys", *Journal of General Internal Medicine*, Vol. 26, No. 5, pp. 498-504.

Fromme, E.K., T. Kenworthy-Heinige and M. Hribar (2011), "Developing an Easy-to-use Tablet Computer Application for Assessing Patient-reported Outcomes in Patients with Cancer", *Support Care Cancer*, Vol. 19, No. 6, pp. 815–822, http://dx.doi.org/10.1007/s00520-010-0905-y.

Fujisawa,R. and N. Klazinga (forthcoming), "Measuring Patient Experiences (PREMs): Progress Made by the OECD and its Member Countries 2006-2015", OECD Publishing, Paris.

Gandhi, S.K. et al. (2001), "Psychometric Evaluation of the 12-item Short-form Health Survey (SF-12) in Osteoarthritis and Rheumatoid Arthritis Clinical Trials", *Clinical Therapeutics*, Vol. 23, No. 7, pp. 1080-1098.

Garratt, A.M., E. Solheim and K. Danielsen (2008), "National and Cross-national Surveys of Patient Experiences: A Structured Rreview", Rapport No. 7-2008, Nasjonalt kunnskapssenter for helsetjenesten (Norwegian Knowledge Centre for the Health Services), Oslo.

Gibbons, E. et al. (2015), "The Outcomes and Experiences Questionnaire: Development and Validation", *Patient Related Outcome Measures*, Vol. 16, No. 6, pp. 179-189, http://dx.doi.org/10.2147/PROM.S82784.

Gonçalves Bradley, D.C. et al. (2015), "Routine Provision of Information on Patient-reported Outcome Measures to Healthcare Providers and Patients in Clinical Practice (Protocol)", The Cochrane Library.

Gordon, M. et al. (2013), "Factors Influencing Health-related Quality of Life After Total Hip Replacement-– A Comparison of Data from the Swedish and Danish Hip Arthroplasty Registers", *BMC Musculoskeletal Disorders*, Vol. 14, No. 316.

Gray, L.C. et al. (2009), "Sharing Clinical Information Across Care Settings: The Birth of an Integrated Assessment System", *BMC Health Services Research*, Vol. 9, No. 71, http://dx.doi.org/10.1186/1472-6963-9-71.

Greene, M.E. et al. (2015), "The EQ-5D-5L Improves on the EQ-5D-3L for Health-related Quality-of-life Assessment in Patients Undergoing Total Hip Arthroplasty", *Clinical Orthopaedics and Related Research*, Vol. 473, No. 11, pp. 3383-3390.

Greenhalgh, J. et al. (2014), "Functionality and Feedback: A Protocol for a Realist Synthesis of the Collation, Interpretation and Utilisation of PROMs Data to Improve Patient Care", *BMJ Open*, Vol. 4, No. 7:e005601.

Haverman, L. et al. (2013), "Effectiveness of a Web-Based Application to Monitor Health-Related Quality of Life", *Pediatrics*, Vol. 31, No. 2, pp. e533-e543, http://dx.doi.org/10.1542/peds.2012-0958.

Hawker, G.A. et al. (2011), "Measures of Adult Pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP)", *Arthritis Care & Research* (Hoboken), Vol. 63, Suppl. 11, pp. S240-252.

Health & Social Care Information Centre (2015), "PROMS Methodologies", available at: http://www.hscic.gov.uk/proms-methodologies (accessed 24/09/2015).

Hearn, J. and I.J. Higginson (1999), "Development and Validation of a Core Outcome Measure for Palliative Care: The Palliative Care Outcome Scale. Palliative Care Core Audit Project Advisory Group", *Quality in Health Care*, Vol. 8, No. 4, pp. 219-227, http://dx.doi.org/10.1136/qshc.8.4.219.

Howell, D. et al. (2015), "Patient-reported Outcomes in Routine Cancer Clinical Practice: A Scoping Review of Use, Impact on Health Outcomes, and Implementation Factors", *Annals of Oncology*, Vol. 6, No. 9, pp. 1846-1858, http://dx.doi.org/10.1093/annonc/mdv181.

Hutchings, A. et al (2014), "Estimating Recruitment Rates for Routine Use of Patient Reported Outcome Measures and the Impact on Provider Comparisons", *BMC Health Services Research*, Vol. 14, No. 66, http://dx.doi.org/10.1186/1472-6963-14-66.

Hutchings, A. et al. (2012), "Factors Associated with Non-response in Routine Use of Patient Reported Outcome Measures After Elective Surgery in England", *Health and Quality of Life Outcomes*, Vol. 10, No. 34, http://dx.doi.org/10.1186/1477-7525-10-34.

ICHOM – International Consortium for Health Outcomes Measurement (2014), "Electronic PROMs: What's the Right Solution for Your Organization?", Cambridge, United States.

ICHOM (2015a), "The Right Thing for Patients: Pioneering Outcomes Measurement in Wales", Cambridge, United States, available at: http://www.ichom.org/others/the-right-thing-for-patients-pioneering-outcomes-measurement-in-wales/ (accessed 03/11/2015).

ICHOM (2015b), "Our Standard Sets", Cambridge, United States, available at: http://www.ichom.org/medical-conditions/ (accessed 09/09/ 2015).

ICHOM (2015c), "Standard Set for Hip & Knee Osteoarthritis", , Cambridge, United States.

InterRAI (2015), "Quality of Life (QoL)", available at: http://interrai.org/quality-of-life.html (accessed 05/10/2015).

Jenkinson, C. et al (1997), "The Parkinson's Disease Questionnaire (PDQ-39): Development and Validation of a Parkinson's Disease Summary Index Score", *Age and Ageing*, Vol. 26, No. 5, pp. 353-357, http://dx.doi.org/10.1093/ageing/26.5.353.

Johnston, B.C. et al. (2013), "Patient-reported Outcomes in Meta-analyses – Part 2: Methods for Improving Interpretability for Decision-makers", *Health and Quality of Life Outcomes*, Vol. 11, No. 1, p. 211.

Kahn, K.L. et al. (2007), "Patient centered Experiences in Breast Cancer: Predicting Long-term Adherence to Tamoxifen Use", *Medical Care*, Vol. 45, No. 5, pp. 431-439, http://dx.doi.org/10.1097/01.mlr.0000257193.10760.7f.

Kosinski, M. et al. (1999), "The SF-36 Health Survey as a Generic Outcome Measure in Clinical Trials of Patients with Osteoarthritis and Rheumatoid Arthritis: Relative Validity of Scales in Relation to Clinical Measures of Arthritis Severity", *Medical Care*, Vol. 37, 5. Suppl., MS23-39.

Le QA (2014), "Probabilistic Mapping of the Health Status Measure SF-12 onto the Health Utility Measure EQ-5D Using the US-population-based Scoring Models", *Quality of Life Research*, Vol. 23, No. 2, pp. 459-466.

McCann, L. et al. (2009), "Patients' Perceptions and Experiences of Using a Mobile Phone-based Advanced Symptom Management System (ASyMS) to Monitor and Manage Chemotherapy Related Toxicity", *European Journal of Cancer Care*, Vol. 18, No. 2, pp. 156-164, http://dx.doi.org/10.1111/j.1365-2354.2008.00938.x.

McHorney, C.A. and J.A. Fleishman (2006), "Assessing and Understanding Measurement Equivalence in Health Outcome Measures. Issues for Further Quantitative and Qualitative Inquiry", *Medical Care*, Vol. 44 (11), Suppl. 3, pp. S205-210.

Murray, D.W. et al. (2007), "The Use of the Oxford Hip and Knee Scores", *Journal of Bone and Joint Surgery*, Vol. 89, No. 8, pp. 1010-1014.

Nag, N. (106), "Development of Indicators to Assess Quality of Care for Prostate Cancer", *European Urology Focus,* Epub ahead of print.

Nilsdotter, A.K. et al. (2003), "Hip Disability and Osteoarthritis Outcome Score (HOOS) – Validity and Responsiveness in Total Hip Replacement", *BMC Musculoskeletal Disorders*, Vol. 4, No. 10.

Nilsson, E., L. Orwelius and M. Kristenson (2016), "Patient-reported Outcomes in the Swedish National Quality Registers", *Journal of Internal Medicine*, Vol. 279, No. 2, pp. 141-153.

NZOA – New Zealand Orthopedic Association (2016), "The New Zealand Joint Registry: Seventeen Year Report – January 1999 to December 2015". http://nzoa.org.nz/system/files/NZJR%2017%20year%20Report.pdf

Obradovic, M., A. Lal and H. Liedgens (2013), "Validity and Responsiveness of EuroQol-5 Dimension (EQ-5D) versus Short Form-6 Dimension (SF-6D) Questionnaire in Chronic Pain", *Health and Quality of Life Outcomes*, Vol. 11, No. 110.

Osborn, R. et al. (2014), "International Survey of Older Adults Finds Shortcomings In Access, Coordination, and Patient-Centered Care", *Health Affairs*, The Commonwealth Fund, Vol. 33, No. 12, pp. 2247-2255, http://dx.doi.org/10.1377/hlthaff.2014.0947.

Oude Voshaar, M.A. et al. (2014), "Linking Physical Function Outcomes in Rheumatology: Performance of a Crosswalk for Converting Health Assessment Questionnaire Scores to Short Form 36 Physical Functioning Scale Scores", *Arthritis Care & Research* (Hoboken), Vol. 66, No. 11, pp. 1754-1758.

Parry, C. et al. (2008), "Assessing the Quality of Transitional Care: Further Applications of the Care Transitions Measure", *Medical Care*, Vol. 46, No. 3, pp. 317-322, http://dx.doi.org/10.1097/MLR.0b013e3181589bdc.

Partridge, T. et al. (2016), "Improving Patient Reported Outcome Measures (PROMs) in Total Knee Replacement by Changing Implant and Preserving the Infrapatella Fatpad: A Quality Improvement Project", *BMJ Quality Improvement Reports*, Vol. 5, No. 1.

Peters, M. et al. (2013a), "Pilot Study of Patient Reported Outcome Measures (PROMs) in Primary Care", Report to the Department of Health, University of Oxford.

Peters, M. et al. (2013b), "Carer Quality of Life and Experiences of Health Services: A Cross-sectional Survey Across Three Neurological Conditions", *Health and Quality of Life Outcomes*, Vol. 11, No. 103, http://dx.doi.org/10.1186/1477-7525-11-103.

Reeve, B.B. et al (2013), "ISOQOL Recommends Minimum Standards for Patient-reported Outcome Measures Used in Patient-centered Outcomes and Comparative Effectiveness Research", *Quality of Life Research*, Vol. 22, No. 8, pp. 1889-1905, http://dx.doi.org/10.1007/s11136-012-0344-y.

Rolfson, O. et al. (2016), "Patient-reported Outcome Measures in Arthroplasty Registries", *Acta Orthopaedica*, Vol. 16, No. 87, Suppl. 1, pp. 3-8.

Rolfson, O. et al. (2011), "Use of Patient-Reported Outcomes in the Context of Different Levels of Data", *Journal of Bone and Joint Surgery*, Vol. 93, Suppl. 3, pp. 66-71, http://dx.doi.org/10.2106/JBJS.K.01021.

Rolfson, O. et al. (2016), "Patient-reported Outcome Measures in Arthroplasty Registries. Report of the Patient-Reported Outcome Measures Working Group of the International Society of Arthroplasty Registries Part II. Recommendations for Selection, Administration, and Analysis", *Acta Orthopaedica*, Vol. 87, Suppl. 1, pp. 9-23.

Sansoni, J. (2016), "Health Outcomes: An Overview from an Australian Perspective", Australian Health Outcomes Collaboration; Australian Health Services Research Institute, University of Wollongong.

Sequist, T.D. et al. (2008), "Quality Monitoring of Physicians: Linking Patients' Experiences of Care to Clinical Quality and Outcomes", *Journal of General Internal Medicine*, Vol. 23, No. 11, pp. 1784-1790, http://dx.doi.org/10.1007/s11606-008-0760-4.

Smith, S.C. et al. (2005), "Measurement of Health-related Quality of Life for People with Dementia: Development of a New Instrument (DEMQOL) and an Evaluation of Current Methodology", *Health Technology Assessment*, Vol. 9, No. 10.

Snowden, A. (2015), "Health Care Quality Report Minneapolis", Minnesota Community Measurement.

Strömqvist, B. and Swedish Society of Spinal Surgeons (2013), "Swespine: The Swedish Spine Register: The 2012 Report", *European Spine Journal*, Vol. 22, No. 4, pp. 953-974, http://dx.doi.org/10.1007/s00586-013-2758-9.

Takeuchi, E.E. et al. (2011), "Impact of Patient-Reported Outcomes in Oncology: A Longitudinal Analysis of Patient-Physician Communication", *Journal of Clinical Oncology*, Vol. 29, No. 21, pp. 2910-2917, http://dx.doi.org/10.1200/JCO.2010.32.2453.

Taylor, R.M. et al. (2015), "Development and Validation of the BRIGHTLIGHT Survey, A Patient-reported Experience Measure for Young People with Cancer", *Health and Quality of Life Outcomes*, Vol. 13, No. 107, http://dx.doi.org/10.1186/s12955-015-0312-7.

Van Noorden, M. et al. (2013), "Routine Outcome Monitoring in Psychiatric Clinical Practice: Background, Overview and Implications for Person-centered Psychiatry", *European Journal for Person Centered Healthcare*, Vol. 1, No. 1, pp. 103-111, http://dx.doi.org/10.5750/ejpch.v1i1.640.

Varagunam, M. et al. (2014), "Impact on Hospital Performance of Introducing Routine Patient Reported Outcome Measures in Surgery", *Journal of Health Services Research & Policy*, Vol. 19, No. 2, pp. 77-84.

ViBIS – Danish Kowledge Center for User Involvement in Health Care (2015), "Patient Reported Outcome (PRO) Data for Quality Improvement of Care Pathways in the Danish Health Care System: Draft Recommendations from Program PRO", ViBIS, Copenhagen.

Ware, J.E. Jr. and C.D. Sherbourne (1992), "The MOS 36-item Short-form Health Survey (SF-36). I. Conceptual Framework and Item Selection", *Medical Care*, Vol. 30, No. 6, pp. 473-483.

Ware, J.E. Jr. et al. (1995), "A. Comparison of Methods for the Scoring and Statistical Analysis of SF-36 Health Profile and Summary Measures: Summary of Results from the Medical Outcomes Study", *Medical Care*, Vol. 33, 4 Suppl., pp. AS264-279.

Wu, A.W., R.E. Jensen and C. Salzberg (2013), "Advances in the Use of Patient Reported Outcome Measures in Electronic Health Records", Johns Hopkins Bloomberg School of Public Health, Baltimore, United States.

Wu, A.W. et al. (2005), "Creating a Crosswalk to Estimate AIDS Clinical Trials Group Quality of Life Scores in a Nationally Representative Sample of Persons in Care for HIV in the United States", *HIV Clinical Trials*, Vol. 6, No. 3, pp. 147-157.

Xian, Y. et al. (2015), "Real World Effectiveness of Warfarin Among Ischemic Stroke Patients with Atrial Fibrillation: Observational Analysis from Patient-Centered Research into Outcomes Stroke Patients Prefer and Effectiveness Research (PROSPER) Study", *British Medical Journal*, Vol. 351:h3786, http://dx.doi.org/http://dx.doi.org/10.1136/bmj.h3786.

Annex A

Reviews assessing effectiveness of PROMs in clinical practice, for quality improvement and performance measurement

Author	Year	Setting	Outcomes	Findings
1. Greenhalgh and Meadows N=13 RCTs	1999	Clinical practice	Patient satisfaction	N=3: 0
			Patient-clinician communication/SDM	N=1: =; N=3: 0
			Screening	N=6: +
			Clinical decision making	N=3: +; N=7: 0
			Health outcomes	N=2: +; N=3: 0
2. Espallargues et al. N=21 RCTs	2000	Clinical practice	Process of care	N=11: +; N=9: 0
			Health outcomes	N=4: + ; N=7: 0
3. Gilbody et al. N=9 RCTs	2001	Depression and anxiety in non-psychiatric setting	Screening	N=4: 0
			Health outcomes	N=4: 0
4. Gilbody et al. N=9 (R)CTs	2002	Mental health in non-psychiatric settings	Patient satisfaction	N=1: +
			Monitoring	N=5: +; N=3: 0
			Health outcomes	N=1: +; N=4: 0
5. Gilbody et al. N=16 (R)CTs	2003	Mental health in primary care	Screening	N=3: 0
			Screening (high-risk patients)	N=2: +
			Clinical decision making	N=2: +; N=7: 0
			Health outcomes	N=7: 0 ; N=1: +
6. Marshall et al. N=38 (35 RCTs, 3 CCTs)	2006	Clinical practice	Patient satisfaction	N=2: +; N=5: 0
			Patient-clinician communication/SDM	N=2: +; N=1: 0
			Screening / Monitoring	N=9: +; N=8: 0
			Health outcomes	N=1: +; N=1: 0
7. Valderas et al. N=34 RCTs	2008	Clinical practice	Screening/diagnosis	N=7:+; N=7: 0
			Patient-clinician communication/SDM	N=3: +;N=4: 0
			Health outcomes	N=8: +; N=9: 0
8. Luckett et al. N=6 RCTs	2009	Oncology	Patient Satisfaction	N=2: +/0
			Clinical decision making	N=1: +
			Health outcomes	N=5: +/0
9. Chen et al. N=27 studies (16 RCTs;9 B/A; 2 OBS)	2013	Oncology	Patient satisfaction	N=13: +; N=3: 0
			Patient-clinician communication/SDM	N=21: +; N=1: 0 ; N=1: -
			Screening	N=15: +; N=1: 0
			Monitoring	N=11: +
			Health outcomes	N=13: +; N=2: 0
			Quality improvement	*No studies found*
			Performance measurement	*No studies found*

Author	Year	Setting	Outcomes	Findings
10. Boyce and Browne 17 RCTs	2013	Clinical practice	Health outcomes	N=1: +; N=6: +/0; N=9: 0
			Quality improvement	*N=1: 0*
11. Boyce et al. 16 studies	2013	Clinical practice	Qualitative research investigating the experiences of healthcare professionals	Barriers and facilitators in 4 themes: (1) PROMs data collection, (2) value of PROMs data, (3) making sense of data, (4) using data to make changes to patient care
12. Howell et al. N=30 studies (7 RCTs; 4 cohort; 5 feasibility; 4 SR; 10 other)	2015	Oncology	Patient satisfaction	N=2 RCTs: 0
			Patient-clinician communication/SDM	N=7 RCTs: +
			Screening	N=2 RCTs: +
			Monitoring	N=1 RCTs: +
			Clinical decision making	N=3 RCTs: +
			Health Outcomes	N=1 RCTs: + ; N=4 RCTs: 0; N=1 RCT: +/0
13. Kendrick et al. N=17 RCTs	2016	Mental Health	Improving clinical management	N= 7 RCTs: 0
			Health outcomes	N=12 RCTs: 0

1. Greenhalgh, J. and K. Meadows (1999), "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", *Journal of Evaluation in Clinical Practice*, Vol. 5, No. 4, pp. 401-416.

2. Espallargues, M., J.M. Valderas and J. Alonso (2000), "Provision of Feedback on Perceived Health Status to Health Care Professionals: A Systematic Review of its Impact", *Medical Care*, Vol. 38, No. 2, pp. 175-186.

3. Gilbody, S.M., A.O. House and T.A. Sheldon (2001), "Routinely Administered Questionnaires for Depression and Anxiety: Systematic Review", *British Medical Journal*, Vol. 322, No. 7283, pp. 406-409.

4. Gilbody, S.M., A.O. House and T.A. Sheldon (2002), "Routine Administration of Health Related Quality of Life (HRQoL) and Needs Assessment Instruments to Improve Psychological Outcome – A Systematic Review", *Psychological Medicine*, Vol. 32, No. 8, pp. 1345-1356.

5. Gilbody, S.M. et al. (2003), "Improving the Detection and Management of Depression in Primary Care", *Quality & Safety in Health Care*, Vol. 12, No. 2, pp. 149-155.

6. Marshall, S., K. Haywood and R. Fitzpatrick (2006), "Impact of Patient-reported Outcome Measures on Routine Practice: A Structured Review", *Journal of Evaluation in Clinical Practice*, Vol. 12, No. 5, pp. 559-568.

7. Valderas, J.M. et al. (2008), "The Impact of Measuring Patient-reported Outcomes in Clinical Practice: A Systematic Review of the Literature", *Quality of Life Research*, Vol. 17, No. 2, pp. 179-193.

8. Luckett, T., P.N. Butow and M.T. King (2009), "Improving Patient Outcomes Through the Routine Use of Patient-reported Data in Cancer Clinics: Future Directions", *Psycho-oncology*, Vol. 18, No. 11, pp. 1129-1138.

9. Chen, J., L. Ou and S.J. Hollis (2013), "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 Services Research.*, Vol. 13, No. 211.

10. Boyce, M.B. and J.P. Browne (2013), "Does Providing Feedback on Patient-reported Outcomes to Healthcare Professionals Result in Better Outcomes for Patients? A Systematic Review", *Quality of Life Research*, Vol. 22, No. 9, pp. 2265-2278.

11. Boyce, M.B., J.P. Browne and J. Greenhalgh (2014), "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 Quality & Safety*, Feb. 6, 2014.

12. Howell, D. et al. (2015), "Patient-reported Outcomes in Routine Cancer Clinical Practice: A Scoping Review of Use, Impact on Health Outcomes, and Implementation Factors", *Annals of Oncology*, Vol. 26, No. 9, pp. 1846-1858.

13. Kendrick; T. et al. (2016), "Routine Use of Patient Reported Outcome Measures (PROMs) for Improving Treatment of Common Mental Health Disorders in Adults", *Cochrane Database of Systematic Reviews*, July 13, 2016, CD011119.

Annex B

Disease-specific PROMs in hip osteoarthritis

Oxford Hip Score (OHS)

The OHS is a 12-tem intervention specific (total hip arthroplasty) outcome measure. It assesses functional ability, daily activities and pain. Items are answered using a five-point Likert scale and the raw scores are added to obtain an overall sum score ranging between 0 and 48 with higher scores being better. OHS has been mapped to the EQ-5D Index and a 0.02 point change in the EQ-5D Index was equivalent to a 1 point change in the OHS.

Hip Osteoarthritis Outcome Scale (HOOS)

HOOS is a hip-specific outcome measure and was constructed by adding items considered important by patients to the WOMAC to get improved validity for those with less severe disease or higher demands of physical function. The HOOS includes five subscales: Pain, Other Symptoms, Function in Daily Living, Function in Sport and Recreation and Hip-related Quality of Life – with in total 40 items. Each item is scored on a five-point Likert Scale. Items are coded from 0 to 4, none to extreme difficulty respectively. Each subscale score is calculated independently. Scores for each subscale are converted to a 0-100 score by calculating mean score of the individual items of each subscale and divide by 4 (the highest possible score for a single answer option).

HOOS-PS

HOOS Physical Function Short form (HOOS PS) is a five-item short version derived from the two HOOS subscales: Function in Daily Living and Sport and Recreation Function. The HOOS PS has been validated for THA. Each item is scored on a five-point Likert Scale. Items are coded from 0 to 4, none to extreme difficulty respectively. The HOOS-PS questionnaire

is scored by summing the raw response (range 0-20) and then using a nomogram to convert the raw score to a true interval score (0-100). HOOS-PS can be scored in two directions, best to worst and worst to best. See next section for important information on scoring directions.

Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)

The WOMAC assesses pain, stiffness, and physical function in patients with hip and / or knee osteoarthritis. The WOMAC consists of 24 items divided into three subscales: Pain (5 items): Stiffness (2 items): Physical function (17 items). The items can be scored with a five-point Likert scale or VAS.

On the Likert Scale version, the scores are summed for items in each subscale, with possible ranges as follows: pain=0-20, stiffness=0-8, physical function=0-68. A total WOMAC score is created by summing the items for all three subscales. The maximum total score is 96 points.

Measurement properties

Alviar provided an overview of measurement properties of PROMs in hip & knee osteoarthritis. Content validity for the HOOS is well established. Content validity is considered intermediate for the WOMAC and OHS, lacking clear documentation of the item selection process. WOMAC, HOOS and OHS were positively rated for agreement, although the OHS also had indeterminate ratings based on several other studies. Responsiveness has been examined in all instruments through various methods, although data clarifying the responsiveness to clinical change and definition of the minimal clinically important change are mostly lacking.

Alviar also compared the contents of patient-reported instruments used in hip and knee arthroplasty rehabilitation with the International Classification of Functioning, Disability and Health (ICF). The HOOS had the widest coverage for body functions. All tools addressed general mobility but do not fully address relevant areas of activity, participation and environment, suggesting limited clinical applicability. Davis evaluated the HOOS-PS in comparing construct validity and responsiveness to the HOOS. The short HOOS-PS represents homogenous short measures of physical functioning with similar construct validity and responsiveness to the 17-item subscale of the HOOS.

Annex C

Typical response rates in patient surveys

A limitation of international comparisons is the variation in response rates across countries. For example, in the 2014 Commonwealth Fund survey of older adults, the response rate was as low as 16% in Norway, where random digit dialing was used, compared with 60% in Switzerland, which used a nationwide population registry (Osborn et al., 2014). These differences raise the possibility of serious bias.

A London School of Hygiene & Tropical Medicine report for the Department of Health in England recommends 80% recruitment and 80% response rates should be sought to reduce the risk of bias (Browne et al., 2007). Response rates expected and achieved will vary depending on the clinical area, with some areas achieving higher response rates than others. Studies to date indicate participation is higher in elective surgical procedures than in other sectors (Table C.1).

Table C.1. Response rates in the NHS England PROMs programme

Procedure	Pre-operative recruitment rate[1]	Post-operative response rate[2]
Hip replacement	78.40%	85.10%
Knee replacement	81%	85.30%
Hernia repair	54.70%	72.90%
Varicose vein surgery	44.70%	64.80%

Source: 1. Hutchings et al. (2014): relates to patients who underwent surgery between October 2009 and September 2010. 2. Hutchings et al. (2012): relates to patients who underwent surgery between April 2009 and March 2010.

The study examining pre-operative recruitment rates found that, while a recruitment rate over 80% is feasible, this was achieved by only a quarter of providers for hip and knee surgery, and by only 2-4% for hernia repair and

varicose vein surgery. Anecdotal evidence suggests that providers that achieve the highest response rates often have one member of staff, such as a senior nurse, take responsibility for ensuring patients are invited to participate. Patients who did not participate were also more likely to be non-white patients and socially disadvantaged (Hutchings et al., 2014).

In the study examining post-operative response rates in the NHS PROMs programme, adjusted odds ratios demonstrated higher levels of non-response in men, patients aged under 55 years, non-white patients, the most socio-economically disadvantaged patients, those who lived alone, those who had previously undergone similar surgery, those who had been assisted when completing their pre-operative questionnaire, and those who had poorer pre-operative health (Hutchings et al., 2012).

A pilot study on the use of PROMs for long-term conditions in primary care indicates that the response rates are likely to be much lower for chronic disease. In a cohort study, an overall response rate of 38.4% was achieved at baseline, ranging from 30% in asthma patients to 50.4% in heart failure patients. The overall response rate to the cohort follow-up one year after baseline was 71.5%. This suggests that despite low initial response rates, most people who participate in PROMs are willing to continue participating in the longer term. In the cohort follow-up, patients with epilepsy and heart failure were less likely to respond than patients with other long-term conditions. There was also a statistically significant difference in the response rate based on age, region and ethnicity (Table C.2). However, there were no statistically significant differences in the response rate at follow-up for gender, time since diagnosis, number of co-morbidities or by practice (Peters et al., 2013a).

Table C.2. Factors significantly related to the number of questionnaires completed at cohort follow-up

		% responders
LTC (p=0.015)	Asthma	72.9
	COPD	71.4
	Diabetes	75.7
	Epilepsy	62.7
	Heart failure	66.2
	Stroke	74.5
Age (years) (p<0.001)	18-24	37.5
	25-34	48.4
	35-44	63.4
	45-54	70.1
	55-64	78.8
	75-84	71.7
	85+	65.2
Region (p=0.007)	London	67.2
	North-West	73.7
Ethnicity (p=0.008)	White	72.9
	Other	63.0

Note: COPD = chronic obstructive pulmonary disease; LTC = long-term condition.

Source: Peters et al. (2013a).

Following the low response rate at cohort baseline of 38.4%, some changes were introduced to the questionnaires, cover letters and information sheets in an attempt to increase response rates for a cross-sectional survey carried out one year later in a sub-sample of practices. A slightly higher response rate of 44% was achieved. It was difficult to determine whether modifications to the survey were responsible for the increased response rate, or other factors. This group of patients were invited to do a one-off survey rather than a repeated cohort survey, to reduce the burden. Additionally, the practices in the second cross-sectional survey were predominantly from the north west, where a higher response rate had been achieved at cohort baseline (Peters et al., 2013a).

The rate of exclusion ranged from 4.1% for diabetes and 5.8% for chronic obstructive pulmonary disease (COPD), to 18.1% for asthma, more than 20% each for heart failure and stroke, up to 46.7% for epilepsy patients. The high rate of exclusion for epilepsy was related to a high proportion of epilepsy patients with learning difficulties. The levels of exclusion were considered within acceptable limits only in COPD and diabetes (Peters et al., 2013a). Consideration needs to be given as to how to maximise participation of these groups. In some cases, proxy reporting may be needed to complete surveys on behalf of patients. However, this should be done with caution, as proxy ratings may not replicate those of patients.

These issues are more challenging for patients with multiple co-morbidities. In the pilot, a high proportion of patients for each long-term condition at baseline reported one or several additional morbidities: 42.8% for asthma, 77.1% for COPD, 76.8% for diabetes, 57.2% for epilepsy, 80.3% for heart failure, and 88.1% for stroke. Patients with more than one of the six long-term conditions were sent a survey for their rarest condition only, as it was considered too burdensome to ask patients to complete multiple PROMs (Peters et al., 2013a).

These issues raise questions about the need for minimum response rates in order to improve generalisability of the findings, validity and international data comparability for both Patient-reported indicators.

Annex D

International Society for Quality of Life Research (ISOQOL) Recommendations for minimum standards for patient-reported outcome measures

1. Conceptual and measurement model: A PROM should have documentation defining and describing the concept(s) included and the intended population(s) for use. In addition, there should be documentation of how the concept(s) are organised into a measurement model, including evidence for the dimensionality of the measure, how items relate to each measured concept, and the relationship among concepts included in the PRO measure.

2. Reliability: The reliability of a PROM should preferably be at or above 0.70 for group-level comparisons, but may be lower if appropriately justified. Reliability can be estimated using a variety of methods including internal consistency reliability, test–retest reliability, or item response theory. Each method should be justified.

3. Validity – 3a. Content validity: A PROM measure should have evidence supporting its content validity, including evidence that patients and experts consider the content of the PROM relevant and comprehensive for the concept, population, and aim of the measurement application. This includes documentation of as follows: 1) qualitative and/or quantitative methods used to solicit and confirm attributes (i.e., concepts measured by the items) of the patient-reported outcome relevant to the measurement application; 2) the characteristics of participants included in the evaluation (e.g., race/ethnicity, culture, age, gender, socio-economic status, literacy level) with an emphasis on similarities or differences with respect to the target population; and 3) justification for the recall period for the measurement application.

3b. Construct validity: A PROM should have evidence supporting its construct validity, including documentation of empirical findings that support predefined hypotheses on the expected associations among measures similar or dissimilar to the measured patient-reported outcome.

3c. Responsiveness: A PROM for use in longitudinal research study should have evidence of responsiveness, including empirical evidence of changes in scores consistent with predefined hypotheses regarding changes in the measured patient-reported outcome in the target population for the research application.

4. Interpretability of scores: A PROM should have documentation to support interpretation of scores, including what low and high scores represent for the measured concept.

5. Translation of the PROM: A PROM measure translated to one or more languages should have documentation of the methods used to translate and evaluate the PROM in each language. Studies should at least include evidence from qualitative methods (e.g., cognitive testing) to evaluate the translations.

6. Patient and investigator burden: A PROM must not be overly burdensome for patients or investigators. The length of the PROM should be considered in the context of other PROMs included in the assessment, the frequency of patient-reported outcome data collection, and the characteristics of the study population. The literacy demand of the items in the PROM should usually be at a 6th grade education level or lower (i.e., 12-year-old or lower). However, it should be appropriately justified for the context of the proposed application.

Source: Reeve, B.B. et al. (2013), "ISOQOL Recommends Minimum Standards for Patient-reported Outcome Measures Used in Patient-centered Outcomes and Comparative Effectiveness Research", *Quality of Life Research*, Vol. 22, No. 8, pp. 1889–1905, http://dx.doi.org/10.1007/s11136-012-0344-y.

Annex E

The Outcomes and Experiences Questionnaire

The Outcomes and Experiences Questionnaire (OEQ) aims to bring together in one eleven-question instrument questions about outcomes and experience of care. It was developed from literature reviews, iterative drafting and discussion within the research group, and cognitive testing with a sample of patients. It has been trialled in the NHS England, but not yet adopted.

The OEQ-O (outcomes) is a summed scale adding the scores for the individual items Q1, Q2, Q3, Q4 and Q11. Scores range from 0 to 20 with a higher score indicative of a better outcome. The OEQ-E (experience) is a summed scale adding the scores for the individual items Q5, Q6, Q7, Q8, Q9 and Q10. Scores range from 0 to 18 with high scores indicating a good experience.

Q1. How helpful has your most recent visit to hospital been in dealing with the problem(s) you came to hospital for?

□Extremely helpful □Very helpful □Helpful □A little helpful □Not at all helpful □Problem(s) completely cured

Q2. How would you now rate the problem(s) you recently came to hospital for?

□Much better □A little better □The same □A little worse □Much worse

Q3. How helpful was your most recent visit to hospital in helping you manage any aspects of the problem(s) that continued after you left hospital?

□Extremely helpful □Very helpful □Helpful □A little helpful □Not at all helpful □No problems remained; problem(s) completely cured

Q4. How would you rate your health now as a result of your hospital visit?

□Much better □A little better □The same □A little worse □Much worse

Q5. When you had important questions to ask staff, did you get answers that you could understand?

□Yes, always □Yes, most of the time □Yes, some of the time □No, never

Q6. How helpful was the information you were given about your treatment and condition at your most recent hospital visit?

□Extremely helpful □Very helpful □Helpful □A little helpful □Not at all helpful
□I was not given information but would have liked some □I did not need any information

Q7. Were you involved as much as you wanted to be in decisions about your care and treatment at your most recent hospital visit?

□As much as I wanted to be □Less than I wanted to be □Not at all although I wanted to be
□Not at all and I did not wish to be □I was more involved than I wanted to be

Q8. How much did hospital staff respond to your individual needs during your most recent hospital visit?

□At all times □Most of the time □Some of the time □None of the time

Q9. Were you able to discuss any worries and fears with staff during your most recent hospital visit?

□As much as I wanted □Most of the time □Some of the time □Not at all, but would have liked to
□I did not have any worries or fears

Q10. Did the different people treating and caring for you work well together to give you the best possible care?

□Yes, always □Yes, most of the time □Yes, some of the time □No never □Don't know

Q11. Overall, how would you rate the outcome of your most recent visit to hospital?

□Excellent □Very good □Good □Fair □Poor

Source: Gibbons E. et al. (2015), "The Outcomes and Experiences Questionnaire: development and validation", *Patient Related Outcome Measures*, Vol. 16, No. 6, pp. 179-189, http://dx.doi.org/10.2147/PROM.S82784.