Article:
Porter, I, Gonçalves-Bradley, D, Ricci-Cabello, I et al. (6 more authors) (2016) Framework and guidance for implementing patient-reported outcomes in clinical practice: evidence, challenges and opportunities. Journal of comparative effectiveness research, 5 (5). pp. 507-519. ISSN 2042-6305

https://doi.org/10.2217/cer-2015-0014

A review on the use of Patient Reported Outcome (PROs) in clinical settings

Porter I[1], Gonçalves-Bradley D[2], Ricci-Cabello R[3], Gibbons C[4,5], Gangannagaripalli J[1], Fitzpatrick R[2], Black N[6], Greenhalgh J[7], Valderas JM[1].

[1] Health Services & Policy Research Group, Exeter Collaboration for Academic Primary Care (APEx) (Person Centred Care), University of Exeter Medical School, University of Exeter, Exeter, United Kingdom

[2] Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom

[3] Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[4] Primary Care Unit, Cambridge Centre for Health Services Research, University of Cambridge

[5] The Psychometrics Centre, Judge Business School, University of Cambridge

[6] Department of Health Services Research and Policy, London School of Hygiene & Tropical Medicine, London, United Kingdom

[7] School of Sociology and Social Policy, University of Leeds, Leeds, United Kingdom

Abstract

Patient reported outcomes (PROs) refer to any report of the status of a patient's health condition that comes directly from the patient. While PROs are a well-developed technology with robust standards in research their use for informing health care decisions is generally poorly understood. In this paper we review relevant examples of their application in health care provision to date, and examine the challenges associated with routinely implementing PROs in clinical settings. We evaluate evidence for their use, and examine key barriers, such as where there is a lack of alignment between interpretation of scores and clinical decision making on the ground. Based on available evidence we have developed a framework of the key requirements for the successful implementation of PROs in clinical practice. We conclude by exploring potential future developments for the use of PROs in clinical practice, such as individualised measurement and efficient administration through the use of computer adaptive tests (CATs).

[Abstract word limit = 120 words, currently 154]

Introduction

Patient Reported Outcome (PROs) measures (also sometimes referred to as PROMs) are health status assessments elicited directly from patients, usually in the form of standardised questionnaires. To date, PROs have been used mainly for research purposes and occasionally also for monitoring the health of a given population, mostly as part of population surveys. In some countries, notably in the National Health Service (NHS) in England, there has been an explicit effort to put them at the centre of the development of a more patient centred, outcomes oriented performance model [1].

Aggregated measurements of PROs for well-defined populations have a potentially huge role to play a driver of patient centred quality improvement. A critical step towards that aim, however, will be whether success is achieved in their use at the level of individual clinical decision making, which currently constitutes an underdeveloped area of implementation. There is therefore a pressing need to evaluate the challenges of using PROs in clinical practice, scrutinising implementation and the evidence supporting their use.

In this paper we will:

- Provide an overview of what PROs are, including the different types of measure and applications.
- Evaluate the research evidence for the use of PROs in clinical practice.
- Review previous experiences of utilising PROs at both the aggregated and the individual level.
- Provide a framework for the successful implementation of PROs in clinical settings.
- Explore potential future developments for the use of PROs in clinical practice.

Overview of PROs

Over recent years there has been a great expansion in the use of PROs, especially for research purposes, yet there have been some confusion as to what the term "PRO" actually refers to. For the purpose of this review we will use the definition provided by the US Food and Drug Administration, which defines PRO as an umbrella term that incorporates "any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else" [2]. The appeal of this terminology is rooted in the fact that it avoids much of the theoretical debate surrounding the conceptualisation and definition of many key related constructs, such as perceived health, health status, (health related) quality of life and well-being, while at the same time emphasising the genuine importance of the individual's own perspective when making an evaluation on their health [3]. In this paper we refer to PROs in relation to overall constructs and PROMs where we refer instruments operationalising the measurement of the construct.

A previous model-based classification system for research and clinical practice has established three guiding concepts for PROMs: construct, population and measurement [3]. Firstly, construct refers to the range of characteristics (traits and states) measured by the instrument along with its measurement object (e.g. symptom status, functional status, health perceptions, health related quality of life). Secondly, the population of a PROM is the range of people for who the instrument is suitable (e.g. age, gender, condition, and culture). Thirdly, measurement incorporates i) metric (the method used to assign numeric values to the responses given by the individuals and the construction of the scores), ii) dimensionality (the number of scores produced for each individual), iii) adaptability (the extent to which the instrument can be tailored to the specific circumstances and preferences of each individual - the degree to which measurement is standardised or individualised).

The majority of PROMs tend to be presented in the form of fully standardised questionnaires with fixed items and response options. The first widely used PROMs included the Sickness Impact Profile [4], the Nottingham Health Profile [5], and the Instrumental Activities of Daily Living [6]. Most widely used measures today include the EuroQOL EQ-5D [7] and the Short Form SF-36 [8]. These were originally classified as generic measures, e.g. applicable across

and within populations, regardless of gender, and disease or condition. However the traditional distinction between generic and specific instruments has been challenged based on the fact that many so called generic measures do not actually cover all major constructs (symptoms, function, health perceptions, health related quality of life), nor are applicable across all age groups (children, adults, elderly) [9].

Less commonly, instruments may also allow patients to modify their content or scoring system to adapt to their own circumstances. These include individualised PROMs, such as the Person Generated Index [10]. Computer-adaptive PROMs, in which the most relevant items are iterative selected and administered by a learning computer algorithm until sufficient information about the patient is collected, also provide an alternative. Individualised and computer-adaptive PROMs will be considered in greater depth later in the review.

The number of PROMs has greatly expanded in the recent years, and the two most comprehensive databases identify a large number of instruments (PROQOLID (711) [11] and BIBLIOPRO (1356) [12]). The co-existence of competing measures aimed at the same populations has triggered the development of standardised systems for the evaluation of the instruments focussing on psychometric properties (validity, reliability, responsiveness, interpretability and cross-cultural adaptation) and ease of use (including administration burden) with the aim of supporting head to head comparisons, such as EMPRO [13] and COSMIN [14, 15].

PROs have most commonly been utilised for research purposes. They have also been utilised, albeit less frequently, for monitoring the health of a stated population, typically through periodic surveys [16, 17]. In clinical practice, PROs allow for objective measurement of the patient's subjective view of their health status to complement other clinical measures. Numerous benefits for using PROs have been cited, such as the screening, diagnosing, monitoring of conditions, along with the promotion of patient centred care [18]. In theory, however, there are many more potential clinical applications (Box 1).

[INSERT BOX 1 SOMEWHERE AROUND HERE]

Each of these potential applications can be considered plausible because PROs are measures of health status. Unlike measures related to processes of care, which are intermediate products of the health system, these are endpoints relevant to the patient. Unlike measures of patients' satisfaction with care, they do not focus on an evaluation of the standards of care, but rather on the impact that care itself has on the health of individuals. PROs have now, however, joined measures of processes of health care, along with patient experience and satisfaction, in the armamentarium of feedback and measurement tools for improving the quality of health care. There have also been barriers to their use in clinical settings, for example, rather than being considered a helpful addition they have often been regarded as disruptive to health care. There has also been a perceived lack of expertise on the part of clinicians with regards to the interpretation and application of such measures.

Evidence for using PROs in clinical practice

A number of systematic reviews have been conducted aiming to evaluate the impact of using PROMs in clinical practice [33, 34, 35, 9, 27] and a wide reaching Cochrane review on the routine provision of PROMs to healthcare providers and patients is currently underway [36]. Other PROM related systematic reviews have focussed on mental health [37, 38]; oncology [39] and palliative care [40] (one with a focus on barriers and facilitators for implementation, and two scoping reviews have concentrated on aspects of routine feedback, including a typology of the diverse ways in which PROMs have been used in mental health services [41], the impact on health outcomes, and implementation factors in routine cancer care [42].

Of those systematic reviews which have taken a broad perspective including all type of conditions, settings and measures [33, 31, 43, 36] the most cited is Valderas et al (2010) which critically appraised the impact of routinely measuring patient-reported outcomes in clinical practice across 34 randomised clinical trials. These reviews have found evidence for impact upon the processes of care, specifically, an increase in the rate of diagnoses and chart notations for the conditions targeted by the interventions. This observation was confirmed for the specific use of PROMs for screening and case finding of depression in another

systematic review, although this significant although limited effect (Relative Risk (RR) 1.38; 95% CI 1.04 to 1.83) was mostly due to the effect of a reduced number of studies in which patients were considered eligible only when scoring above a certain severity threshold. Similarly, there was also a positive effect on the advice and education provided by the healthcare professionals. However, no changes were observed in terms of medication and similarly, no impact was observed on the management of depression.

The impact of the use of PROMs on outcomes of care remained inconclusive for the vast majority of the systematic review except for the specific use of PROMs for feedback on treatment outcome in mental health specialist settings: a significant positive short-term effect on mental health was observed which did not prevail in the long run (effect size d = 0.10; 95% CI 0.01–0.19).

Valderas [9] identified a total of 36 endpoints for the randomised controlled trials (RCTs) included in the systematic review, which appears to reiterate the lack of consensus amongst researchers in relation to how interventions should work and thus what constitutes a relevant indicator when using PROMs in clinical practice. In addition, these systematic reviews have consistently observed methodological limitations with regards to design and analysis. Critically, many interventions did not consider providing training on how to interpret the information collected by the PROMs. More importantly, many studies demonstrated an incomplete understanding of the mechanisms of by which the intervention operates [9, 18, 44]. In this way, a considerable methodological limitation is that they have attempted to establish the impact of the use PROs on distal primary outcomes, such as improvement on health status, without understanding impact upon proximal outcomes, for instance communication between patients and the healthcare professionals [45, 46].

The potential effect of the use of PROMs on health outcomes is crucially mediated by the modification of the behaviour of both patients and professionals [9, 30]. ‘Patient centredness’ is a critical component in this model, and while all PROMs are self-reports of the patient’s health status, only some of them explicitly focus the evaluation on those areas that are most relevant for each patient. Individualised patient reported outcome measures (iPROMs) are concerned with areas which are directly nominated by patients themselves [3], research

conducted in the UK has highlighted the need to incorporate individualised assessment alongside standard measures [44].

Lessons learned from using PROs at the aggregate level

The extent to which PROs have been implemented in clinical settings globally has, to date, been somewhat varied. PROMs are currently at the centre of an extensive programme for the orientation of the NHS in England, the largest and oldest single-payer healthcare system in the world, towards a performance model based on health outcomes. This marks the most ambitious programme implementing PROMs to date globally, the first time that all patients undergoing specific interventions have been invited to complete measures before and after specific treatments [Link to website]. Given the scale of the NHS programme a detailed examination of implementation in clinical practice in England, and lessons that can be learned from this, is particularly relevant.

The NHS PROMs programme has been operating since 2009. Under this mandatory programme PROMs (the EQ-5D alongside condition-specific instruments), have been collected for all NHS patients undergoing one of four elective surgical procedures (hip replacement, knee replacement, varicose vein, and groin hernia surgery), both before and after surgery [22]. The PROMs Programme has focused on the collection of the aggregation of patient scores for comparison of providers rather than to inform individual patient management and clinical decision making. Information is collected at an individual level while feedback takes place at a group level, while such feedback has limitations, the resources involved in gathering such information in the first place are extensive, and many lessons can potentially be learned from this to inform clinical decision making. A substantial body of knowledge on the routine collection and feedback of these measurements is developing as a result of its evaluation, particularly in terms of rationale, feasibility, and impact [23-31].

[INSERT BOX 2 SOMEWHERE AROUND HERE]

Policy makers now possess evidence about many aspects of the use of PROs in order to drive quality improvement. Firstly, a weak association is observed between PROMs and patient reported experience measures (PREMs) [23]. Secondly, it is feasible to organise data collection and feedback on a national scale, with modest recruitment rates (45%-68% depending on the intervention) with a high post-operative response (65%-85%) [24, 25]. Thirdly, the choice of metric derived from PROMs makes a substantial difference to the results, such as the proportion of providers defined as 'poor' [24].

Although the PROMs programme has been positive in the sense that many participating patients reported valuing their experience, there was, however, no discernible impact on patient outcomes over its first three years [26]. This may be because providers only started receiving feedback in the third year of the programme, in addition the presentation of feedback was, at the time, not of high quality, and there was little room for improvement for the selected procedures. However, the evidence for using feedback on PROMs for quality improvement at a group level is severely limited. A systematic review [27] was only able to identify one study with methodological limitations that, not surprisingly, did not observe a positive impact [28].

The underlying assumption is that measurement of PROs, along with clear feedback and dissemination of this information, will stimulate and incentivise health professionals and ultimately health trusts to provide better care [30]. A key challenge is how to bridge the gap between the information that is provided at the aggregate level with changes in care delivery based on individual patients [1, 31]. So far, these two agendas have developed relatively independently of each other, without joined-up thinking about how they might come together. For example, it will be essential to maintain high response rates for aggregate use of PROMs to monitor quality; this is much more likely if there is engagement by health professionals and patients. More importantly it is difficult to anticipate how clinicians will be able to identify opportunities for improvement of care as measured against the standard of PROM scores unless these scores are part of their daily experience of managing patients. The expansion of the clinical applications of PROMs will be pivotal to the success of the whole PROMs programme. A realist synthesis of the literature is underway [32] to help identify

users' aims and the theories underlying their expectations and will put these observations in an appropriate context.

A framework for implementing PROs in clinical practice

Using available evidence from systematic reviews and from qualitative research with patients and health professionals [43, 47, 48], along with specific guidelines for implementing outcome measures in clinical practice [49-51], and the evidence from available experiences, we have developed a framework of key requirements for the successful implementation of PROs in clinical practice (Box 2).

[INSERT BOX 3 SOMEWHERE AROUND HERE]

Based on this, six key issues have been identified that need to be considered when implementing PROs in clinical practice: the characteristics of the PRO itself, the purpose of using the measure, the feedback system, the setting, specific additional support to the implementation, as well as targeting patients and health professionals that are most likely to benefit from it. Specific requirements include training for clinicians on the administration and interpretation of PROMs; the use of existing information systems for feedback to clinicians, that is frequent and timely; and the provision of information that is linked to specific actions, among others [9, 18, 52].

We present a practical example of how these requirements can be operationalised, based on a study evaluating the use of PROs for children with advanced cancer in the USA (Box 4). This randomised controlled trial, conducted at three large paediatric cancer centres, comprehensively considers the issues of implementing PROMs in clinical practice with lessons that can be applied elsewhere. In this study, the use of PROs was used to support the routine monitoring of children and teens undergoing treatment, and allowed clinicians to systematically track symptoms and health related quality of life over time. The design of the study incorporated the key issues we have identified above (and in Box 2), in particular special consideration was given to feedback for both families and providers, with printed reports and

emails provided immediately after survey completion. Families and providers were also offered training in how to interpret the results. The study authors also acknowledged potential limitations to their feedback intervention, in that it may have lacked leverage required to affect clinical behaviour and subsequent referrals. The limitation of feedback linked to treatment decision making is observed in comparable studies [ref] and constitutes and ongoing barrier to the take up of PROs in clinical practice.

[INSERT BOX 4 SOMEWHERE AROUND HERE]

Our framework for the implementation of PROs in clinical practice also provides insight into the ultimately unsuccessful implementation of PROs for the assessment of depression as part of the Quality and Outcomes Framework (QOF), an incentive scheme for Primary care services covering the vast majority of the population of England (Box 5). In particular, neither training nor specific indications were given on which measure to choose and the rationale for a choice, when or how to introduce the measure, when to collect the score, or how to introduce it within the flow of a clinical appointment, and scarce information was provided regarding the instruments themselves [54]. The feedback system was also considered poor, with the scores not being available in a timely fashion, lack of time for integrating the information provided by PROs with other clinical information, and irregular feedback. The only elements that were considered were to select instruments tailored to the setting and to link PRO scores (depression severity) to specific action (treatment options), through explicit reference to the National Institute for Health and Care Excellence (NICE) guideline for the management of the condition [55] (NICE is a public body providing national guidance and advice to improve health and social care in England). As such, the framework, as outlined, provides a means of learning from the lessons of the past, indicating ways PROs can be more successfully implemented in clinical practice in the future.

[INSERT BOX 5 SOMEWHERE AROUND HERE]

Future developments

The potential of individualised PROs

A key challenge facing health care systems across the world concerns how to promote patient centred care in the context of the increasing demands of patients managing multiple, long-term conditions. It is estimated that between a quarter and half of primary care patients have more than one health problem [61, 62]. Even though clinically complex patients have an increased use of healthcare services than other members of the population, they face worse health outcomes [63, 64]. The use of iPROMs, where the patient can explicitly prioritise which conditions or symptoms to address, offers the potential of aligning the priorities of patients and professionals.

iPROMs are measures for which individual respondents are actively involved in the specification of content and/or scoring of each item [65]. Such measures empower patients to evaluate the relevance of different aspects of their health conditions [66], either by selecting the specific domains, weighing their importance, or both [67, 3]. By tailoring the questionnaire to the person's own perspective, iPROMs have a higher likelihood of detecting issues that may be relevant in clinical practice, as well as being responsive to the individual aspects of health related quality of life [68]. iPROMs are more likely to be able to detect a change in patient's health status than standardised tools alone (in turn, among standardised tools condition specific instruments will tend to prove more responsive than more generic ones). They may be particularly useful for patients with multiple conditions in helping them to convey to practitioners which aspects of their health they would like to be prioritised.

Despite their ability to convey personal preferences and idiosyncratic aspects from the patient's perspective, there are some disadvantages of using iPROMs, both from a structural and a methodological perspective. Structural disadvantages include the burden associated with administration, at the researcher, clinical and patient level [67]. Methodological issues include the absence of a common metric, which raises concerns regarding analyses and comparison of results (not only between patients but also for the same patient across time), and the possibility of successively obtaining floor effects if the items generated by the patient are too idiosyncratic [68]. Moreover, distinct processes are used for item and content generation, resulting in wide-ranging levels of individualisation [68]. This results in measures with different degrees of individualisation, ranging from those individualised at the content

level, where the respondent can nominate the items (e.g., the Goal Attainment Scaling) to those where the items are provided to the respondent, who is then able to rate the importance and relevance of each item, thus being individualised at the scoring level (e.g., the Audit of Diabetes Dependent Quality of Life [69]. Given these issues it is vital to ensure that iPROMs are methodologically rigorous and fit for purpose. Currently none of the major systems for the standardised evaluation of PROs considers degrees and quality of individualisation [9, 15, 70]. There is, therefore, a pressing need to establish, through expert consensus, rigorous evaluation standards for assessing the methodological quality of iPROMs, along with an acceptable definition of such measures.

The administration of different combinations of standardised and individualised PROMs could provide an alternative to the current, compartmentalised remit of health provision, allowing for multiple conditions to be addressed within a single health care consultation and help to drive forwards patient centred prioritisation. But the question still remains about whether the same PRO can be used for both purposes.

Item Banks and computer adaptive tests

As we move further into the information age, the greater availability of 'smart' computing solutions alongside increasing application of advanced psychometric models to health care PROMs grants a number of new and exciting opportunities to enhance PRO use in clinical practice.

In the field of psychometrics, 'modern' measurement theories are increasingly being utilised to create questionnaires that are both shorter and more accurate than those developed using 'classical' methodologies [58, 71]. These include item response theories [72] and the Rasch model [73], are being increasingly adopted in health services research and have recently been used to develop PROs which measure wide-ranging phenomena including pain, depression, fatigue, and the impact of multimorbidity [74-77]. However, in other fields where questionnaires are used to capture 'high stakes' information, such as educational testing and psychometric assessment for employers, they have been considered the industry standard for decades [78].

One key attribute of item response theory is the ability to calibrate precise item parameters (the degree of the underlying phenomena they measure), rather than simply applying an ordinal integer value to a response. In practice, this means that the validated scales may be administered using any number or combination of items; rather than it always being necessary to administer the entire scale to each respondent. This had led to the increasing development of large 'banks' of psychometrically calibrated questionnaire items, known as item banks [79].

One of advantage of item banks is their ability to capitalise on 'intelligent' computerised questionnaire administration algorithms, especially computer adaptive tests (CATs); which can improve the quality of measurement whilst administering fewer items than a standard paper-based questionnaire [80]. Computer adaptive testing improves measurement by selecting the most appropriate items for each patient, based on their previous responses. For example, when administering a functional impairment questionnaire using CATs, the computer would be unlikely to administer an item that represents a very high level of functional impairment (*e.g.* "I have problems getting dressed") once it had decided (following pervious item responses) that the respondent probably had very little impairment (for example, if they responded positively to a statement about exhausting exercise). Questionnaires using CAT usually take a fraction of the time to complete and, because of the rigorous item bank development, can be even more reliable, valid and sensitive to change than compared to their paper-based counterparts [81]. Despite the complexity of the underlying CAT algorithms, their administration is straightforward and has been shown to be acceptable and accurate across diverse populations, including older adults [82] and children [83].

Taken at face value, the increased precision and brevity of CATs may appear to be their most appealing feature in busy clinical environments. But there are other significant benefits to using this technique in clinical practice. The ability of CATs to accurately 'target' questionnaire items means that patients may be spared from responding to items which are either plainly obvious or potentially distressing; such as an item asking a patient in the very early stages of

a progressive degenerative disease if they are experiencing the high levels of impairment that are typical of later disease stages.

The computerised nature of CAT grants a number of other significant benefits for PRO administration. These include the immediate provision of feedback, which can be automatically tailored to the individual responses of the patient. More practical advantages provided by computerised administration include no missing data and automatic scoring, inputting and storing. These issues have all been shown to be important barriers to the uptake of PROs in clinical practice [84].

Future perspectives

The coming years have exciting potential for the use of PROs in clinical practice, and we predict a number of ways in which the field will evolve over the next decade. As technology plays an ever greater part in society and healthcare systems, we anticipate that computerised devices incorporating CAT, such as tablets, will be used for the increased collection and feedback of PROs, making effective use of both standardised and individualised measures, incorporating personalised feedback for both clinicians and patients. We also anticipate such application will be implemented in an increasingly longitudinal fashion, for PRO scores to become part of clinical reviews over time, linked to changes in patient outcomes, activation and goal setting. As part of this it will be important that clinical staff are trained in the understanding and application of PROMs, to avoid mistakes of the past, which will involve developing expertise in how to interpret measures and feed this information back to patients in order to inform treatment. At present there is a lack of link up with PROs and clinical decision making, as to how interpretation of measures can directly influence clinical decision making at the “bed-side”. This constitutes and area that needs to be addressed in coming years, a more comprehensive framework needs to be developed in order to provide greater guidance to clinicians as to how confident they can be in measured values, and how interpretation of such values can be linked to a range of appropriate treatment options.

Conclusion/Discussion

PROs have a role to play in clinical practice and the evidence of this is becoming stronger. However, there is currently more compelling evidence for impact upon the processes of care, than on outcomes of care. A key barrier for implementation continues to be lack of alignment with current clinical processes, particularly in terms of insufficient linkage between interpretation of scores and clinical decision making on the ground.

Based on available evidence we have developed a framework of the key requirements for the successful implementation of PROs in clinical practice, encompassing the instrument itself, purpose of the PRO, the feedback system, setting, specific additional support to implementation, as well as characteristics of patients and health professionals involved (along with training for clinicians on the administration and interpretation of PROs). In the future this framework could be further developed into a "bed-side" tool kit providing clinicians with confidence in values, how to interpret them, and link them to treatment options (in a way that is responsive to the needs of the patient).

We are now starting to see effective of administration and instant feedback of PROs through computerised devices. The next logical step in this is to develop enhanced relevance to the each patient. Most previous efforts have focussed on the use of fully standardised measures thereby limiting opportunities to tailor measurement to individual preferences. To this end individualised measurement and efficient administration through the use of CAT represent very promising developments for further uptake of PROs in clinical practice, administering measures in a way that is appropriate and responsive to each patient and different aspects of their health conditions. Crucially CATs offer a stream lined approach which can reduce burden compared to the time required to complete PROMs through traditional means, this in itself offers a key advantage to stretched health care systems going forwards.

References

[1] Black N. Patient reported outcome measures could help transform healthcare. BMJ 2013;346:f167.

[2] US Food and Drug Administration: Guidance for Industry Patient-reported outcome measures: Use in medical product development to support labeling claims. US: Department of Health and Human Services Food and Drug Administration; December, 2009. Available from:http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf

[3] Valderas, J.M. and J. Alonso, Patient reported outcome measures: A model-based classification system for research and clinical practice. Quality of Life Research, 2008. 17(9): p. 1125-1135.

[4] Bergner, M., Bobbitt, R. A., et al. (1981). The Sickness Impact Profile: development and final revision of a health status measure. Medical care 19(8): 787-805.

[5] Hunt SM, McKenna SP, McEwen J et al. The Nottingham health Profile: subjective health status and medical consultations. Social Sciences and Medicine. 1981;15A:221-229.

[6] Powel Lawton M, Brody E. Assessment of older people: self-maintaining and instrumental activities of daily living. The Gerontologist 1969;9(3):179-186.

[7] Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med. 2001 Jul;33 (5): 337-43.

[8] Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey:construction of scales and preliminary tests of reliability and validity. Med Care. 1996 Mar;34(3):220-33.

[9] Valderas JM, Kotzeva A, Espallargues M, et al. The impact of measuring patient-reported outcomes in clinical practice: A systematic review of the literature. Qual Life Res 2008;17:179-93.

[10] Ruta DA, Garratt AM, Leng M, et al. A new approach to the measurement of quality of life. The patient-generated index. Med Care. 1994;32:1109–1126.

[11] Emery MP, Perrier LL, Acquadro C: Patient-Reported Outcome and Quality of Life Instruments Database (PROQOLID): frequently asked questions. Health Qual Life Outcomes 2005, 3:12.

[12] Garcia-Duran Huet, Mireya, et al. "BiblioPRO: online library of PRO instruments in Spanish." QUALITY OF LIFE RESEARCH. VOL. 19. VAN GODEWIJCKSTRAAT 30, 3311 GZ DORDRECHT, NETHERLANDS: SPRINGER, 2010.

[13] Valderas JM, Ferrer M, Mendívil J, Garin O, Rajmil L, Herdman M, Alonso J; Scientific Committee on "Patient-Reported Outcomes" of the IRYSS Network. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. Value Health. 2008 Jul-Aug;11(4):700-8.

[14] Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HCW. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. BMC Medical Research Methodology 2006;6:2.

[15] Mokkink, L.B., et al., The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. Journal of clinical epidemiology, 2010. (63.7): 737-745.

[16] National Health Interview Survey 2014. Survey description. Centers for Disease Control and Prevention. 2015. Available at: http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm

[17] Black N, Burke L, Forrest CB, Ravens Sieberer UH, Ahmed S, **Valderas JM**, Bartlett SJ, Alonso J. Patient-reported outcomes: pathways to better health, better services, and better societies. Qual Life Res. 2015 Nov 13

[18] Greenhalgh J. The applications of PROs in clinical practice: What are they, do they work, and why? Qual Life Res 2009;18:115-23.

[19] Department of Health. A National Service Framework for Mental Health, 1999. www.gov.uk/government/uploads/system/uploads/attachment_data/file/198051/National_Service_Framework_for_Mental_Health.pdf

[20] Department of Health. Services for people with depression: Report of a CSAG Committee chaired by Professor Chris Thompson. London, 2000.

[21] Roland M. Linking physicians' pay to the quality of care–a major experiment in the United kingdom. N Engl J Med. 2004; 351(14):1448-1454.

[22] Department of Health. Guidance on the routine collection of Patient Reported Outcome Measures (PROMs), 2009. http://www.mstrust.org.uk/competencies/downloads/NHS-PROMS.pdf.

[23] Black N, Varagunam M, Hutchings A. Relationship between patient reported experience (PREMs) and patient reported outcomes (PROMs) in elective surgery. BMJ Qual Safety 2014;23:534-42.

[24] Hutchings A, Grosse Frie K, Neuburger J, van der Meulen J, Black N. Late response to patient reported outcome questionnaires was associated with worse outcome. J Clin Epidemiol 2013;66:218-225

[25] Hutchings A, Neuburger J, van der Meulen J, Black N. Estimating recruitment rates for routine use of patient reported outcome measures and the impact on provider comparisons. BMC Health Serv Res 2014;14:66 doi:10.1186/1472-6963-14-66

[26] Varagunam M, Hutchings A, Neuburger J, Black N. Impact on hospital performance of introducing routine patient reported outcome measures (PROMs) in surgery. J Health Serv Res Pol 2014;19:77-84

[27] Boyce MB, Browne JP. Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients? A systematic review. Qual Life Res. 2013 Nov;22(9):2265-78.

[28] Weingarten SR, Kim CS, Stone EG, et al. Can peer-comparison feedback improve patient functional status? American Journal of Managed Care 2000; 6(1), 35–39.

[29] Peters M, Crocker H, Dummett S, Jenkinson C, Doll H, Fitzpatrick R. Change in health status in long-term conditions over a one year period: a cohort survey using patient-reported outcome measures. Health Qual Life Outcomes. 2014 Aug 12;12:123. doi: 10.1186/s12955-014-0123-2.

[30] Greenhalgh J, Long AF, Flynn R. The use of patient reported outcome measures in routine clinical practice: Lack of impact or lack of theory? Soc Sci Med 2005;60:833-43.

[31] Valderas JM, Ferrer M, Mendivil J, et al. Development and testing of standards for individualization of patient reported outcome (PRO) measures using EMPRO. Qual Life Res 2012;21(S1):43

[32] Greenhalgh J, et al. "Functionality and feedback: a protocol for a realist synthesis of the collation, interpretation and utilisation of PROMs data to improve patient care." BMJ open 4.7 (2014): e005601.

[33] Espallargues M, Valderas JM, Alonso J. Provision of feedback on perceived health status to health care professionals: A systematic review of its impact. Medical Care 2000; 38, 175–186.

[34] Greenhalgh, J., & Meadows, K. (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, 5, 401–416.

[35] Marshall, S., Haywood, K., & Fitzpatrick, R. (2006). Impact of patient reported outcome measures on routine practice: a structured review. Journal of Evaluation in Clinical Practice, 12(5), 559–568.

[36] Gonçalves Bradley DC, Gibbons C, Ricci-Cabello I, Bobrovitz NJH, Gibbons EJ, Kotzeva A, Alonso J, Fitzpatrick R, Bower P, van der Wees PJ, Rajmil L, Roberts NW, Taylor RS, Greenhalgh J, Porter I, Valderas JM. Routine provision of information on patient-reported

outcome measures to healthcare providers and patients in clinical practice. Cochrane Database of Systematic Reviews 2015, Issue 3. Art. No.: CD011589. DOI: 10.1002/14651858.CD011589

[37] Gilbody, S. M., House, A. O., & Sheldon, T. A. (2001). Routine administered questionnaires for depression and anxiety: Systematic review. British Medical Journal, 322, 406–409.

[38] Knaup C, Koesters M, Schoefer D, et al: Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. British Journal of Psychiatry 195:15–22, 2009

[39] Alsaleh, Khalid. "Routine administration of standardized questionnaires that assess aspects of patients' quality of life in medical oncology clinics: A systematic review." Journal of the Egyptian National Cancer Institute 25.2 (2013): 63-70.

[40] Etkind, et al. "Capture, Transfer, and Feedback of Patient-Centered Outcomes Data in Palliative Care Populations: Does It Make a Difference? A Systematic Review." Journal of pain and symptom management (2014)

[41] Krägeloh CU, Czuba KJ, Billington DR, Kersten P, Siegert RJ. Using Feedback From Patient-Reported Outcome Measures in Mental Health Services: A Scoping Study and Typology. Psych Serv 2014 (3):224-241. http://dx.doi.org/10.1176/appi.ps.201400141

[42] Howell, D., et al. "Patient-Reported Outcomes in Routine Cancer Clinical Practice: A Scoping Review of Use, Impact on Health Outcomes, and Implementation Factors." Annals of Oncology (2015): mdv181.

[43] Gonçalves DC, Gibbons EJ, Fitzpatrick R, Valderas JM. A systematic review and narrative review of qualitative studies about the use of PROs in clinical practice. Qual Life Res 2012;21:25

[44] Valderas JM, Espallargues M, Kotzeva A, Alonso J. Assessing the impact of routinely measuring patient-reported outcomes in clinical practice: critical appraisal of 34 randomized clinical trials. International Society of Quality of Life Research. London, UK. November 2010. Published in: Qual Life Res 17(Suppl 1), 2010.

[45] Whooley, M.A., B. Stone, and K. Soghikian, Randomised trial of case-finding for depression in elderly primary care patients. Journal of General Internal Medicine, 2000. 15(5): p. 293-300.

[46] Bergus, G.R., et al., The limited effect of screening for depressive symptoms with the PHQ-9 in rural family practices. Journal of Rural Health, 2005. 21(4): p. 303-309.

[47] Dowrick C, Leydon GM, McBride A, et al. Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: Qualitative study. BMJ 2009;338:b663.

[48] Leydon GM, Dowrick CF, McBride AS, et al. Questionnaire severity measures for depression: a threat to the doctor-patient relationship? Br J Gen Pract 2011;61(583):117-23

[49] Snyder CF, Aaronson NK, Choucair AK, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. Qual Life Res 2012;21:1305-14.

[50] Valderas JM, Alonso J, Guyatt GH. Measuring patient-reported outcomes: moving from clinical trials into clinical practice. Med J Aust. 2008 Jul 21;189(2):93-4.

[51] Valderas JM, Alonso J, Prieto L, Espallargues M, Castells X. Content-based interpretation aids for health-related quality of life measures in clinical practice. An example for the visual function index (VF-14). Qual Life Res. 2004 Feb;13(1):35-44

[52] Valderas JM, Rue M, Guyatt G, Alonso J; Systematic Use of Quality of Life Measures in the Clinical Practice Working Group. The impact of the VF-14 index, a perceived visual function measure, in the routine management of cataract patients. Qual Life Res. 2005 Sep;14(7):1743-53.

[53] Velikova G, Booth L, Smith AB, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: A randomised controlled trial. *J Clin Oncol* 2004;22:714-24.

[54] British Medical Association - NHS Employers. Revisions to the GMS Contract 2006/07. Delivering Investment in General Practice. In: Department of Health, ed., 2006. http://www.nhsemployers.org/SiteCollectionDocuments/Revisions_to_the_GMS_contract_-_full_CD_120209.pdf

[55] NICE. Depression: The treatment and management of depression in adults. In: National Institute for Health and Clinical Excellence, ed., 2009. http://www.nice.org.uk/nicemedia/live/12329/45888/45888.pdf

[56] Malpass A, Shaw A, Kessler D, Sharp D. Concordance between PHQ-9 scores and patients' experiences of depression: A mixed methods study. Br J Gen Pract 2010;60:e231-8.

[57] Mitchell C, Dwyer R, Hagan T, Mathers N. Impact of the QOF and the NICE guideline in the diagnosis and management of depression: a qualitative study. Br J Gen Pract 2011;61:e279-89.

[58] Prieto L, Alonso J, Lamarca R. Classical Test Theory versus Rasch analysis for quality of life questionnaire reduction. Health Qual Life Outcomes. 2003;1(1):27. doi:10.1186/1477-7525-1-27.

[59] Toop L. The QOF, NICE, and depression: a clumsy mechanism that undermines clinical judgment. Br J Gen Pract 2011;61:432-3.

[60] British Medical Association - NHS Employers. 2013/14 general medical services (GMS) contract quality and outcomes framework (QOF). Guidance for GMS contract 2013/14. http://www.nhsemployers.org/Aboutus/Publications/Pages/qof-2013-14.aspx

[61] Barnett, K., et al., Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet, 2012. 380(9836): p. 37-43.

[62] Salisbury, C., et al., Epidemiology and impact of multimorbidity in primary care: A retrospective cohort study. British Journal of General Practice, 2011. 61(582): p. e12-e21.

[63] Fortin, M., et al., Relationship between multimorbidity and health-related quality of life of patients in primary care. Qual Life Res, 2006. 15(1): p. 83-91.

[64] Smith, S.M., et al., Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. Cochrane Database Syst Rev, 2012. 4: p. CD006560.

[65] Browne, J. http://www.kingsfund.org.uk/sites/files/kf/media/Professor%20John%20Browne%20-%20Individualised%20Patient%20Outcomes.pdf 2014

[66] Garratt, A.N., et al., Quality of life measurement: Bibliographic study of patient assessed health outcome measures. British Medical Journal, 2002. 324(7351): 1417-1419.

[67] Dijkers, M.P., Individualisation in quality of life measurement: Instruments and approaches. Archives of Physical Medicine and Rehabilitation, 2003. 84(4): S3-S14.

[68] Jolles, B.M., R. Buchbinder, and D.E. Beaton, A study compared nine patient-specific indices for musculoskeletal disorders. Journal of Clinical Epidemiology, 2005. 58(8): 791-801.

[69] Bradley, C., et al., The development of an individualised questionnaire measure of perceived impact of diabetes on quality of life: The ADDQoL. Quality of Life Research, 1999. 8(1-2): 79-91.

[70] Fitzpatrick, R., et al., Evaluating patient-based outcome measures for use in clinical trials. Health Technology Assessment, 1998. 2 (14).

[71] Wright B. Raw Scores Are Not Linear Measures: : Rasch vs. Classical Test Theory Comparison. Rasch Meas Trans. 1992;6(1).

[72] Lord FM. Applications of Item Response Theory to Practical Testing Problems. Routledge; 1980. https://books.google.com/books?hl=en&lr=&id=_qu8YGkdTvwC&pgis=1. Accessed May 11, 2015.

[73] Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen: Danish Institute for Educational Research; 1960.

[74] Lai J-S, Dineen K, Reeve BB, et al. An item response theory-based pain item bank can enhance measurement precision. J Pain Symptom Manage. 2005;30 (3):278-288. doi:10.1016/j.jpainsymman.2005.03.009.

[75] Gibbons CJ, Mills RJ, Thornton EW, et al. Development of a patient reported outcome measure for fatigue in motor neurone disease: the Neurological Fatigue Index (NFI-MND). Health Qual Life Outcomes. 2011;9(1):101. doi:10.1186/1477-7525-9-101.

[76] Gibbons C, Kenning C, Coventry P, Bee P. Development of a multimorbidity illness perceptions scale (MULTIPleS). PLoS One. 2013;8(12):e81852.

[77] Forkmann T, Boecker M, Norra C, et al. Development of an item bank for the assessment of depression in persons with mental illnesses and physical diseases using Rasch analysis. Rehabil Psychol. 2009;54(2):186-197. doi:10.1037/a0015612.

[78] Hambleton, Ronald K. Principles and selected applications of item response theory. Linn, Robert L. (Ed), (1989). Educational measurement (3rd ed.). The American Council on Education/Macmillan series on higher education., (pp. 147-200). New York, NY, England: Macmillan Publishing Co, Inc; American Council on Education, xiv, 610 pp.

[79] Holland P, Wainer H. Differential Item Functioning. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.; 2012.

[80] Lai J, Cella D, Chang C-H, Bode RK, Heinemann AW. Item banking to improve, shorten and computerize self-reported fatigue: An illustration of steps to create a core item bank from the FACIT-Fatigue Scale. Qual Life Res. 12(5):485-501. doi:10.1023/A:1025014509626.

[81] Haley SM, Raczek AE, Coster WJ, Dumas HM, Fragala-Pinkham MA. Assessing mobility in children using a computer adaptive testing version of the pediatric evaluation of disability inventory. Arch Phys Med Rehabil. 2005;86(5):932-939. doi:10.1016/j.apmr.2004.10.032.

[82] Jette AM, Haley SM, Ni P, Olarsch S, Moed R. Creating a Computer Adaptive Test Version of the Late-Life Function and Disability Instrument. Journals Gerontol Ser A Biol Sci Med Sci. 2008;63(11):1246-1256. doi:10.1093/gerona/63.11.1246.

[83] Haley SM, Fragala-Pinkham MA, Dumas HM, et al. Evaluation of an item bank for a computerized adaptive test of activity in children with cerebral palsy. Phys Ther. 2009;89(6):589-600. doi:10.2522/ptj.20090007.

[84] Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. BMJ. 2010;340(jan18_1):c186. doi:10.1136/bmj.c186.

Box 1. Potential clinical applications of PROs

- Supporting decision making in the diagnostic process:
  - Screening
  - Diagnosis
- Informing risk stratification and prognosis (identification of vulnerable patients and patients "at risk")
- Supporting prioritisation and goal setting
- Supporting decision making in indication for treatment (medical/surgical)
- Facilitating monitoring of
  - General health status
  - Response to treatment/management
- Facilitating communication
  - Between patients and health professionals
  - Within teams and between professionals: consistent use along the care pathway

Box 2. Potential applications of feedback on PROs for quality improvement.

- Within a single service:

-Evaluation of performance of the service as a whole along the time and for audit

-Comparing outcomes between different user groups

-Contributing to the evaluation of individual clinicians' practice and/or for comparing performance across clinicians in the same service

- Across multiple services:

-Evaluation of performance of the system as a whole along the time

-Comparing performance of individual services

-Comparing outcomes between different user groups

Box 3. Framework of key issues and requirements for the successful implementation of PROs in clinical practice

- Clinical activity:
  - Screening
  - Diagnosis
  - Risk stratification and prognosis
  - Prioritisation and goal setting
  - Indication for treatment (medical/surgical)
  - Monitoring of
    - General health status
    - Response to treatment/management
  - Facilitating communication
    - Between patients and health professionals
    - Within teams and between professionals (along care pathways)
- PROM instrument:
  - Simple (low burden of administration)
  - Valid (in particular, face validity)
  - Responsive
  - Interpretable: explicit link between PROMs outputs and clinical activity
  - Tailored to the particular setting and purpose
- Feedback system:
  - Integrated in clinical information systems
  - Structured, with explicit interpretation of the individual scores
  - Targeting both all relevant health care professionals and patients
  - Frequent and timely
  - Supported by training of patients and health professionals on the interpretation of scores and outputs
- Setting:
  - Questionnaire friendly environment
- Support to the implementation:
  - Training
  - Additional resources

- In addition, the effectiveness of the implementation can be maximised by targeting specific groups most likely to benefit from this approach of
    - Patients: male, previously unknown to the clinician, with perceived borderline status of measurement in relation to the clinical activity
    - Health professionals: low familiarity/degree of confidence in the clinical activity

Box 4. Use of feedback of PROs aiming to improve the Care of Children with Advanced Cancer

This is the first RCT designed to test a PRO feedback intervention among children and teens with advanced cancer. Wolfe and colleagues [x] assessed 339 children for eligibility, with 104 allocated to either the intervention arm of the study or the control group. The control group completed questionnaires using tablet computers no more than once a week: Paediatric Quality of Life and Evaluation of Symptoms Technology (PQ) survey consisting of age and respondent-adapted versions of the Memorial Symptom Assessment Scale (MSAS), Paediatric Quality of Life Inventory 4.0 Generic Core Scales (PedsQL4.0), and an overall Sickness question. Oncologists and families received printed reports summarising PROs; e-mails were sent to oncologists and subspecialists when predetermined scores were exceeded. The control group completed the questionnaires but no feedback was provided. The feedback intervention had two components: PQ reports and PQ e-mails. PQ reports were printed reports that were given to providers before the visit and to families immediately after survey completion. PQ reports (Data Supplement) consisted of bar plots of PedsQL4.0 and MSAS symptom scores from current and four prior administrations, a summary highlighting changes since the last report, and a list of available resources for symptom control (for families) or generic pain management recommendations (for providers). Training on how to interpret PQ reports was offered to families at enrolment and annually to providers. Primary outcomes included linear trends of MSAS, PedsQL4.0 total and subscale scores, and Sickness scores during 20 weeks of follow-up, along with child, parent, and provider satisfaction with PQ feedback. Although routine feedback of PROs did not significantly affect the child's symptoms or health-related quality of life

(HRQL), changes were in expected directions and improvements observed in emotional HRQL through exploratory analyses were encouraging. Importantly, children, parents, and providers valued PRO feedback. Post hoc subgroup analyses found larger and statistically significant improvements on the HRQL emotional domain scores and the sickness question for patients who survived longer than 20 weeks.

Box 5. PROs in the Quality and Outcomes Framework (QOF)

In addition to PROMs being implemented as part of the NHS PROMs programme, PROMs have also been implemented in Primary Care as part of the Quality and Outcomes Framework (QOF), a voluntary annual reward and incentive scheme in Primary Care. The programme was, once again, considered unsuccessful. Incentives for the assessment of the severity of depression at diagnosis and at follow up using a standardised PROMs were introduced in the QOF in 2006 [54]. Minor changes were introduced in the course of the subsequent years. Other than being directed to three specific scales validated for using in primary care (Patient Health Questionnaire [PHQ], Beck Depression Inventory Second Edition [BDI-II], and Hospital and Anxiety Depression Scale [HADS]), [55], GPs received no specific guidance regarding the implementation of the data collection and feedback system.

The indicators were quickly adopted by the practices, as evidenced by data provided by the Health and Social Care Information Centre (Appendix). In 2006/2007 two out of three practices achieved the maximum amount of points, a proportion that increased to four out of five by 2011/2012. Although these figure may indicate a degree of success, qualitative research with patients and health professionals suggested that the use of PROs was not having the desired impact on either group.

Patients overall favoured the use of PROs, which they considered a useful supplement to clinical judgement that could aid the GP in the diagnosis, management, and monitoring of depression [47, 56]. They also noted their usefulness in providing feedback and raising awareness of symptoms of depression [47, 56]. Of particular interest, patients perceived

the administration of these instruments as confirmation of their problems and as evidence that their GP was taking their problems seriously [47].

The perceptions of clinicians, however, were substantially different. Multiple concerns were raised in relation to the applicability, administration, and interpretation of the PROMs. Lack of specific training [57] and simply not knowing when and how to administer the measures were major problems. Faced with a lack of guidance on the questionnaires might be used in clinical practice and the challenge of fitting their administration into a ten minute consultation GPs often introduced significant changes to established administration procedures, affecting where and how the measure was completed, and the number of questions asked [57] and compromising the validity of the scores [58]. It was also not always clear to GPs how to interpret the scores [48]. The PROMs were often not considered as an integral part of assessment and diagnosis, and their administration was felt to disrupt the normal flow of the consultation, [48, 57], while adding an additional burden to the workload and roles of GPs and practice nurses [59]. Perhaps not surprisingly GPs concerns on the use of these PROMs have resulted in their use no longer being specifically incentivised as part of QOF [60].