

Title: "Patient-Reported Outcomes: pathways to better health, better services and better societies"

Black N¹, Burke L², Forrest CB³, Ravens Sieberer UH⁴, Ahmed S⁵, Valderas JM⁶, Bartlett SJ⁷, Alonso J⁸

^{1,2,3}Co-first authors; ^{7,8}Co-senior authors

¹ Department of Health Services Research and Policy, London School of Hygiene & Tropical Medicine, London WC1H 9SH, United Kingdom.

² Founder, LORA Group, LLC Royal Oak, MD, United States

³ Department of Pediatrics, The Children's Hospital of Philadelphia, Pennsylvania, United States.

⁴ Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, United States.

⁵ Faculty of Medicine, McGill University, appointments in the School of Physical and Occupational Therapy, the University Health Center's Division of Clinical Epidemiology, Clinical and Health Informatics, and the Centre de recherche interdisciplinaire en réadaptation, Montreal, QC, Canada.

⁶ Health Services & Policy Research, Exeter Collaboration for Primary Care, University of Exeter Medical School, University of Exeter, Exeter, United Kingdom.

⁷ Faculty of Medicine, McGill University, Montreal, QC, Canada.

⁸ Health Services Research Unit, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain; CIBER Epidemiología y Salud Pública (CIBERESP), Spain; Pompeu Fabra University (UPF), Barcelona, Spain.

[4171 words]

Abstract

While the use of PROs in research is well established, many challenges lie ahead as their use is extended to other applications. There is consensus that health outcome evaluations that include PROs along with clinician reported outcomes and administrative data are necessary to inform clinical and policy decisions. The initiatives presented in this paper underline evolving recognition that PROs play a unique role in adding the patient perspective along side clinical (e.g. blood pressure) and organisational (e.g. admission rates) indicators for evaluating the effects of new products, selecting treatments, evaluating quality of care and monitoring the health of the population. In this paper, we first explore the use of PRO measures to support drug approval and labeling claims. We critically evaluate the evidence and challenges associated with using PRO measures to improve health-care delivery at individual and population levels. We further discuss the challenges associated with selecting from the abundance of measures available, opportunities afforded by agreeing on common metrics for constructs of interest, and the importance of establishing an evidence base that supports integrating PRO measures across the healthcare system to improve outcomes. We conclude that the integration of PROs as a key end point within individual patient care, healthcare organization and program performance evaluations, and population surveillance will be essential for evaluating whether increased healthcare expenditure is translating into better health outcomes.

[224 words]

Key Messages

- While the use of PROs in research is well established, many challenges lie ahead as their use is extended to a wide range of additional applications, such as the regulatory approval process of drugs and medical devices, the monitoring of the health of the populations and the quality of the health services, and supporting decision making in clinical settings.
- A 3-step roadmap developed by the US FDA for PRO measurement in clinical trials demands: 1. understanding the disease or condition; 2. conceptualizing treatment benefits; and 3. adequate selection and development of PRO.
- The England NHS Programme has provided evidence that the systematic collection of PRO is feasible at such a large scale, and that it provides relevant information about health services delivery.
- Several general population studies in child and adolescent in Europe have informed the development and monitoring of policies and strategies at regional, national and international levels.
- There is an increasing number of PROs and other health indicators assessing the same underlying construct but they are most frequently not scaled to the same metric. Interoperability of PROs is possible and necessary for comparability of results obtained in different settings, such as health surveys and medical records.

1. Introduction

Patient Reported Outcomes (PROs) are measurements of health as reported by patients about themselves and include (but are not limited to) symptoms, functioning, health perceptions and health related quality of life [1]. Several thousand PRO measures are in existence and new ones are being developed every day [2]. PRO measures were originally developed, in part, to aid in evaluating and comparing the effectiveness of healthcare interventions. While the use of PROs in research is well established, many challenges lie ahead as their use is extended to other applications.

Real time evaluation of patients' preferences, health, well-being, and behaviour has the potential to inform clinical care and contribute to comparative effectiveness research embedded within real-world care delivery [3,4]. However, it remains unclear how to best aggregate and integrate patient-level PRO data to inform quality improvement and population health surveillance. Standards of how PRO measures are selected, collected, interpreted, reported and merged with other clinical or administrative datasets are needed to ensure that results are valid and meaningful for clinical care and policy decision-making.

Global organizations are addressing these challenges and translating the wealth of expertise in PROs into areas as different as quality improvement, clinical care, and population health surveillance. Initiatives such as the PROMs programme in the England [5], the integration of Patient Reported Outcomes Measurement Information System (NIH-PROMIS) measures in electronic clinical records [6], the routine use of PRO measures in geriatric and mental health settings [7,8], and the inclusion of generic measures in national health surveys are testimony of the vital role of PROs in assessing health outcomes.

In this paper, we first explore the use of PRO measures to support drug approval and labeling claims. We critically evaluate the evidence and challenges associated with using PRO measures to improve health-care delivery at individual and population levels. We conclude with reflections on the challenges associated with selecting from the abundance of measures available, opportunities afforded by agreeing on common metrics for constructs

of interest, and the importance of establishing an evidence base that supports integrating PRO measures across the healthcare system to improve outcomes.

Use of PRO Measures to Support the Approval of Drugs, Labels and Promotional Claims (Laurie Burke)

PRO measures play a vital role in the evaluation of medical intervention benefit and risk. In the US, for instance, any medical product claim of improvements in symptoms or daily functioning must be based on patient-reported assessments that are *well-defined and reliable*, and are used to define study endpoints in *adequate and well-controlled studies*. Even when survival is the primary objective of a study, PRO assessments are necessary to characterize symptoms and daily functioning in those whose life may be extended by treatment. From a regulatory perspective, the only situations in which PROs are not relevant are when patients cannot respond for themselves (e.g., in young children) or when biomarkers or clinical assessment are the only outcomes of disease activity (e.g., early after diagnosis for some hematologic disorders).

Clinical trial measurement science continues to evolve since the entry of the efficacy requirement for medical product marketing in the US in 1962 [9]. PRO measure validation is a continuous process with evidence continuing to be compiled throughout the instrument's lifespan. Planning PRO measurement, well in advance of when it is needed (i.e., with the phase 3 study protocol) can help mitigate a number of challenges as described below.

First, a certain minimum set of information needs to be available before the phase 3 protocol. This includes evidence of **content validity** (i.e., that the score represents the outcome/concept of interest in the context of use), **test-retest reliability** in stable patients similar to those who will be enrolled in the clinical trial at baseline, and basic cross-sectional **construct validity** that demonstrates the relationship of the score to other trial endpoints. Longitudinal measurement properties are also important but, if necessary, can be explored using blinded clinical trial data, with some risk to the trial outcome (e.g., if ability to detect change is inadequate). The interpretation of trial findings is dependent on establishing

thresholds for meaningful change (i.e., defining responders). Decision-makers must interpret study results to determine the importance of the treatment effect in the context of their patient situations or population characteristics and the value of the treatment in comparison to other treatments.

To facilitate this process, the US Food and Drug Administration (FDA) has published a three-step "roadmap" to evaluate and develop PRO measures to support medical product labelling and advertising [10] (Table 1). The first step, "*Understanding the Disease or Condition*", often overlooked in its entirety, represents the process of gathering relevant information that determines the clinical trial context of use. Most importantly for PRO measures, it includes the process of gathering a full understanding of the patient/caregiver perspectives on the definition of treatment benefit, benefit-risk trade-offs, and the impact of the disease on daily life and function. The second step, "*Conceptualizing Treatment Benefit*", consolidates the relevant background information from the previous step to determine the specific population to be studied and the appropriate study design. This step includes identification of the full set of outcomes, PRO and other, to define study endpoints (primary and secondary) that need to be tested in that population to fully understand treatment benefit in light of the possible risks that will be discovered in the clinical trial. Each outcome needs to be identified in terms of the concept represented and its context of use including relative order among all study endpoints included in the statistical analysis.

<Table 1>

With the context of use and the concept of interest fully identified, the third step, "*Selecting/Developing Outcome Measures*," outlines the process of measure identification or development. Measure qualification activities at European Medical Agency (EMA) and FDA, combined with advancements in measurement science, are improving the likelihood that an existing measure will be found. If none exists, however, instrument modification or new measure development is required. Because measures with content validity in the planned clinical trial context are needed before the onset of phase 3, early consideration of the roadmap is hoped to ensure the availability of adequate information for patients and other decision-makers about the impact of treatment on symptoms and functioning at the time of product availability.

Health Care Services: Improving Clinical Practice and Quality (Nick Black)

The routine use of PROs linked to clinical practice has been driven by two principal motives: comparison of outcomes (usually hospitals) and improving the clinical management of individual patients. While the latter has been the main motivation in Sweden and the USA, use in England has centered around comparing the performance of providers [11]. The hope has been that PROs can increase productivity by avoiding unnecessary treatments and improving quality through service redesign and patient choice.

Since 2009, use of PROs has been mandatory in the 250,000 patients undergoing one of four elective surgical operations funded by the English National Health Service (NHS) under the NHS PROMs programme. All patients are invited to complete a pre-operative questionnaire which queries socio-demographic characteristics, comorbidities, and includes a condition-specific PRO and a generic PRO (the EQ-5D). Either three (hernia repair, varicose vein surgery) or six (hip and knee replacement) months after surgery patients are mailed another questionnaire that includes the two PROs plus questions on complications and single transitional items.

Five years on, much has been learnt about the feasibility of the approach and about some methodological properties of PROs used in this way. First, although recruitment rates are modest (68% for hip and knee replacement; 45% for the minor procedures), this does not appear to bias the comparisons of hospitals' outcome as these are adjusted for the factors associated with non-recruitment (over 75 years of age; deprived; non-white) [12]. Second, post-operative response is good (85% for hip and knee replacement; 65-75% for minor procedures) [13]. Third, despite some surgeons' concern that patients may confuse outcome (effectiveness) with their experience of the way care was delivered (humanity), there is only a weak association ($r = 0.2$) between PRO scores and patient reported experience measures (PREMs) [14]. In other words, PROs and PREMs measure different aspects and patients can and do distinguish these domains of quality. Fourth, the choice of metric derived from PROs makes a difference to the proportion of providers defined as 'poor'. For example, for hip

replacement the mean risk-adjusted change in the disease-specific PRO classified 25% were more than 2 standard deviations (SD) from the overall mean. However, using the proportion of patients achieving a minimally important difference identified resulted in only 12% being classified as outliers [15]. If the goal is to avoid missing an outlier, then the former is best; if avoidance of mistakenly labelling a provider as an outlier is the priority, then the latter is the optimal approach.

Experience with PROs has also shed light on aspects of health services delivery. First, although it was expected that increased surgical rates in an area would be associated with a decrease in the mean pre-operative severity of patients, no significant association was found (Figure 1) [16]. This suggests that the decision to operate is still largely determined by surgeons rather than reflecting patients' self-reported need. Second, while the association between hospital volume and safety (such as operative mortality) is well-established, no such association has been found with effectiveness (i.e., risk-adjusted change in PROM scores) [17]. Third, little difference in outcomes has been observed between individual surgeons, though more outliers (i.e., > 3 SD from the expected) are detected than when post-operative mortality is used as the outcome [18]. And finally, despite widely held beliefs that competition results in better quality, no association has been seen between the amount of competition in a locality and mean risk-adjusted change in PRO scores [19].

<Figure 1>

The NHS PROMs Programme, however, had no discernible impact on patient selection or outcomes over its first three years [20]. This may be for three reasons: 1) providers only started receiving feedback in the third year of the programme; 2) the quality of presentation of feedback was rather poor (indigestible spreadsheets); or 3) for these four procedures, there is little room for improvement – all hospitals and surgeons provide good quality care already. Greater differences with an associated greater scope for improvement are more likely with emergency surgical operations.

One of the main research challenges ahead is a need to clarify expectations as to the routine use of PROs. A realist synthesis of the literature is underway [21] to help identify users' aims and the theories underling their expectations. There are also methodological challenges to

address regarding the routine use of PROs in more taxing clinical areas such as long term conditions (with no before and after measurement feasible/appropriate), emergency conditions (with no before measurement possible) and dementia (where cognitive dysfunction limits the extent to which patients can contribute).

Other challenges concerns policy and practice. There is a need to bring together the two principal reasons for using PROs – clinical management and provider performance assessment. Ensuring that future use combines both aims will enhance the quality of the data collected, patient and clinician engagement, and the potential benefits for policy-makers [22]. These aims will be facilitated by the adoption of new data collection techniques that minimise data collection costs and offer timely feedback. Both clinical management and provider assessment would benefit from better presentation of output to result in actionable feedback.

Well-being in Populations: The Case for Monitoring Child Mental Health (Ulrike Ravens-Sieberer)

Mental health constitutes "the foundation for well-being and effective functioning for an individual and for a community" [23]. A substantial proportion of mental health problems in adults originate early in life [24]. Although the health of children and adolescents has been continuously improving, children and adolescents today are more likely to experience social, emotional and behavioural difficulties resulting in range of disorders including depression, anxiety, disruptive disorders and eating disorders [25]. Monitoring children's health and well-being is a promising strategy that can aid in early detection of hidden or manifest mental health problems [26,27]. Here, we make the case that PROs be used as an indicator of population health, albeit with acknowledgement that the use of positive mental health and well-being indicators has not deserved extensive attention [28].

Three European studies have developed and implemented a range of indicators to assess child and adolescent mental health and well-being. The *Health Behaviour in School-aged Children (HBSC) Study*, in collaboration with the WHO Regional Office for Europe, has been

collecting health information for 25 years in over 40 countries in Europe and North America [26,29,30]. The *European KIDSCREEN project* developed and implemented a standardized screening instrument for health-related quality of life (HRQL) in children and adolescents in 13 European countries (Figure 2) [28]. The *Flash Eurobarometer* collected parent's views on the mental health and well-being of their children in 27 EU member states in 2008 [31].

<Figure 2>

Such studies provide robust data on a wide range of child and adolescent health issues that can help inform the development of policies and strategies [32]. They also provide researchers, clinicians and policy makers with the knowledge necessary to evaluate and advance development of public health policy and practice at regional, national and international levels [33]. The WHO/HBSC Forum Series is a good example of how research data can be leveraged for policy making in WHO European member countries. In Forum meetings, HBSC data are used to promote discussion among international partners and facilitate the translation of research findings into effective policy-making and practice (www.hbsc.org). The products of these meetings are a synthesis report and policy statement, along with additional materials. Past reports have addressed healthy eating habits and physical activity (2006), social cohesion for mental well-being (2007), and socio-environmentally determined health inequities (2009).

Current research is illness-oriented and, as previously noted, the development of positive mental health indicators is largely still in its infancy. Too often, mental health is misconceptualized as the absence of mental disorders, as evidenced by the fact that existing measures used to assess population mental health are primarily "needs driven" and focus on "illness" and distress [34], such as delinquency, suicide, depression, rather than "wellness". Lack of comparable measures, reporting methods, and cultural differences in perceptions (i.e., whether mental disorders are transient or not, and whether and how they require treatment in different countries and communities) also impede the reliable assessment of current trends.

The RICHE Roadmap report on gaps and needs in child health research in Europe clearly identified several factors including: 1) the need for early detection (screening) and

monitoring; 2) the use of age- and culturally-appropriate instruments; and 3) the need to place a stronger focus on younger age groups. At present, measures are generally adolescent-focused and young children are generally underrepresented in international data sources. Monitoring of children's mental health, along with screening tools to detect groups at higher risk for poor mental health is an important act of prevention and a promising strategy to detect negative developments [35].

To date the effect of early childhood risk factors remains understudied, and research into factors associated with the development of mental disorders and the means by which they interact is necessary to develop effective primary prevention strategies and interventions. To obtain comprehensive and comparable information on the current health of children and adolescents today, a European health survey for children in all age groups (from infants to teenagers) needs to be established [35]. There is also a need for more European comparative longitudinal studies in specific health areas, to conduct meaningful cross-cultural/cross country comparisons through the use of validated and routinely implemented indicators, and to further develop age-related indicators in specific areas [35].

Monitoring the well-being of our youth is clearly important. Health indicators are a bridge between health policy and scientific information [36]. Effective mental health indicators can facilitate international comparisons and provide valuable information about the well-being of children, leading to better identification of vulnerable groups and areas needing support. Measuring well-being internationally can lead to greater acknowledgement of the importance of well-being in young people everywhere [35].

Linking Population Health, Clinical Research, and Clinical Practice PROs

Applications: The Call for a Common Metric (Chris Forrest)

Several recent developments in the United States highlight the important role of PROs in population health, large-scale clinical research, and clinical practice. These include the nation's health objectives (i.e., Healthy People 2020 [37]), the national clinical research network-of-networks, funded by the Patient Centered Outcomes Research Institute (PCORI)

and called PCORnet [38], and an Institute of Medicine (IOM) report on the social and behavioral domains that should be included in all electronic health records [39] (IOM 2014).

Healthy People is the U.S. government's decennial national objectives for population health improvement and reduction of disparities [37]. Four domains provide the core set of measures for evaluating progress: 1) determinants of health; 2) general health status; 3) disparities; and 4) HRQL and well-being. Determinants of health subsume behavior, health services, social factors, and policy. General health status includes conventional indicators such as life expectancy, years of potential life lost, limitations of activity, and chronic disease prevalence. Disparities cover assessments of the health indicators by such factors as socioeconomic status, sex, sexual identity, race or ethnicity, age, disability, and geographic location. Lastly, for the first time, Healthy People added a focus on health-related quality of life. After reviewing measures of self-reported health, well-being, and participation, Healthy People 2020 selected the PROMIS® adult global health measure, which includes physical health scale and mental health scales [40], for monitoring objectives. The scales are turned into indicators by dichotomizing them at scores reflective of good or better self-reported physical or mental health. Healthy People 2020 has set targets for improving national adult self-reported health, is collecting data to enable monitoring of disparities in these indicators and is now considering extending these indicators to children and youth.

In 2014, PCORI announced the launch of PCORnet, a national network of networks designed to conduct patient-centered outcomes research [38]. PCORnet includes 29 research networks that are harmonizing their data to a common data model. An example of one of the networks is PEDSnet, which includes 8 large pediatric health systems and 3 disease-specific networks [41]. It is expected that PCORnet, when fully mature and conducting research, will include somewhere between 50-100 million Americans. The PCORnet common data model includes electronic health record data and PROs. It is envisioned that research networks will embed the collection of PROs into routine clinical care, and these measures will be used to describe study samples, to control for differences between groups, and to assess primary and secondary end-points in observational studies and clinical trials.

The IOM issued a report in November, 2014 proposing a concise set of self-reported social and behavioural determinants (SBD) of health that should be included in all patients' electronic health records [39]. It is well recognized that social and behavioural factors are as important, and in some cases more important, than biomedical factors in determining health and health risk. Yet, they are not captured during medical encounters in any systematic way, which effectively precludes their use to improve patient care, leading to calls to create approaches for routinely characterizing patients' psychosocial environment on a periodic basis [42]. Recognizing the opportunity to address this need, nine federal and foundation sponsors asked the IOM to assemble a committee of social science, clinical, and informatics experts to recommend a set of SBDs for inclusion in electronic health records. Placing priority on measures with the greatest clinical usefulness and feasibility for capture within the clinical workflow, the committee developed a parsimonious panel of complementary measures that when given in its entirety includes 24 questions, which would take about 5-10 minutes to complete before or during a medical encounter. The measures cover socio-demographic, psychological, behavioral, social relationship, and community factor dimensions, providing a comprehensive assessment of a patient's psychosocial environment.

As exciting as each of these new developments is, they all suffer from the challenge of measure selection. We lack a set of [universally accepted measures for a comprehensive measurement of PROs, particularly addressing psychosocial constructs and quality of life](#). Some may argue that this is a strength of our field, as it reflects our ingenuity and creativity as scientists. However, imagine that the research networks in PCORnet choose to measure social isolation with measure A, while the clinical practices that adopt the IOM recommendations choose to measure the same concept with measure B. Although the two measures may assess the same underlying construct, they are not scaled to the same metric, rendering the data obtained in these two applications non-comparable. In informatics terms, the two measurement systems are not interoperable, which prevents facile integration of findings from research studies into clinical practice, meta-analyses using the two sources of data, and data sharing.

The proliferation of measures for the same construct creates a Tower of PRO-Babel that may be unnecessary. We argue that quality of life researchers should work together to agree on common metrics for constructs of interest. This is akin to saying that all temperature will be measured on the Kelvin scale, which allows devices to obtain assessments using Celsius or Fahrenheit scales assuming that they are then converted to the Kelvin scale (Table 2). In other words, use the measurement device of your choosing, but represent the data on the same scale. The PROMIS measurement system provides a set of common metrics to which PRO instruments that assess comparable constructs can be scaled. Each PROMIS measure undergoes item response theory calibration and national norming of the calibrated scale to a mean of 50 and standard deviation of 10. Any item or scale that measures the same domain as a given PROMIS measure can be transformed to the PROMIS T-score, allowing for users to collect data using the measure of their choice, but rendering it on a common metric. However, to date, this has been done for a minority of instruments [43]. The convergence of electronic health record proliferation and integration of PROs into clinical practice, clinical research, and population health creates a new imperative for tearing down the Tower of PRO-Babel and erecting in its place a common language with common metrics.

<Table 2>

Discussion

There is consensus that health outcome evaluations that include PROs along with clinician reported outcomes and administrative data are necessary to inform clinical and policy decisions. The integration of PROs as a key end point within individual patient care, healthcare organization and program performance evaluations, and population surveillance will be essential for evaluating whether increased healthcare expenditure is translating into better health outcomes.

The numerous PROs already developed offer an indication of the growth of the field, but in areas where there are multiple choices, this can lead to greater uncertainty and create additional barriers for to uptake. We have identified gaps in PRO applications that can inform strategies to manage the selection, application, and interpretation of PROs.

Regardless of the setting or application, similar challenges exist in the widespread use and application of PROs.

First, the use of particular measures to evaluate domains of interest, when measures are applied to new populations or contexts, should be supported by an evaluation of their metric properties in these new circumstances. This resource-intensive activity is often overlooked when a measure is applied to a new patient population or context. Second, guidance is needed for the selection of measures that are reliable, valid and responsive to change[44, 45].

Standardisation of metrics is needed to facilitate comparisons. Finally, interpretation of scores is also critical as this will influence clinical and policy decisions, and in turn, the relevance of PROs for improving access and quality of care[19,45]. Identifying effective formats for the feedback of PRO data to stakeholders, and consensus on actions is a key intermediary step to achieving improvements in quality of care.

These considerations underline several areas for future research and changes to healthcare policy. This includes facilitating the implementation of PROs with as regard to measure selection, ease of use, and feedback of recommendations that are actionable. Further evaluation is also needed to determine how the choice of metrics influences the conclusions and changes in practice and to validate PRO data with comparisons to other sources of data.

As the application of PROs in clinical care and health policy is relatively new, further research is needed to evaluate whether using them makes a difference. Studies addressing these issues are currently underway [21,47].

Finally, further mapping is needed to define governance and infrastructure to support collection and feedback of data so that PROs can inform a learning health care ecosystem [48]. Careful planning of how data from different sources will be merged and accessed, including consensus across electronic health records (HER) vendors and administrative databases, and a common coding system for PRO domains and items is also needed. The infrastructure and established processes also must ensure that the most vulnerable (e.g. elderly, those with learning difficulties or dementia) who are just as likely to benefit from healthcare improvements, are not excluded when PRO data are collected. Such an infrastructure will facilitate the application of PROs not only for clinical care and health

policy, but also to generate the evidence in support of the belief that incorporating the patient voice through PROs can help make things better.

Acknowledgements

The work was presented in the Plenary session "Patient-Reported Outcomes: contributing to better services and better societies" at the 2014 Annual Meeting of the International Society for Quality of Life Research (ISOQOL) in Berlin, Germany. We would like to thank ISOQOL for their support in the organization of the teleconferences leading to the preparation of the session and to a very engaged audience for their questions, which have helped shape the discussion of several points in this manuscript.

Compliance with Ethical Standards

Disclosure of Potential Conflicts of Interest. The authors do not report any other conflict or potential conflict of interest. Jose M Valderas was supported by an NIHR Clinician Scientist Award (NIHR/CS/010/024).

Ethical approval: This article does not contain any studies with animals performed by any of the authors.

Summary of author contributions

This plenary was held Thursday 16 October 2014 in Berlin, Germany. JA and SJB conceived the session and provided steer for the preparation of the plenary and manuscript and contributed equally to the manuscript. SA and JV organized the session, and served as co-chairs. SA developed summaries of the panel discussion. NB, LB, CBF, and UHRS contributed equally to the content of the session and to the manuscript. NB, LB, CBF were featured speakers. SA and JV equally co-led the development of manuscript based on drafted sections by NB, LB CBF and UHRS summarizing their talks. All authors helped revise and improve initial drafts and read and approved the final manuscript.

References

1. Valderas JM, Alonso J. "Patient reported outcome measures: a model-based classification system for research and clinical practice." *Quality of Life Research* 17.9 (2008): 1125-1135.
2. Quality of Life Instruments Database-PROQOLID [computer program]. Lyon, France: Mapi Research Institute; 2010.
3. Greenhalgh, J., The applications of PROs in clinical practice: What are they, do they work, and why? *Quality of Life Research*, 2009. 18(1): p. 115-123
4. Valderas JM, Kotzeva A, Espallargues M, Guyatt G, Ferrans C, Revicki D, Halyard M, Parada A, Alonso J. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Qual Life Res* 2008;17(2):179-93
5. Health & Social Care Information Centre. Patient Reported Outcome Measures. Available at: <http://www.hscic.gov.uk/PROs> [accessed 19/01/2015]
6. Christopher B CB Forrest. Digitization of patient-reported outcomes. *Value in Health* 16(4):459
7. American Psychiatric Association. DSM 5. American Psychiatric Association, 2013.
8. Kane, Robert L., et al. *Essentials of clinical geriatrics*. New York: McGraw-Hill, 1994.
9. Act of October 10, 1962 (Drug Amendments Act of 1962), Public Law 87-781, 76 STAT 780, which amended the Federal Food, Drug, and Cosmetic Act to assure the safety, effectiveness, and reliability of drugs, authorize standardization of drug names, and clarify and strengthen existing inspection authority, 10/10/1962.
10. US Food and Drug Administration. Roadmap to Patient-Focused Outcome Measurement in Clinical Trials. Available at: <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm370177.htm> [accessed 19/01/2015]
11. Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013;346:19-21
12. 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
13. Hutchings A, Neuburger J, Gross Frie K, van der Meulen J, Black N. Factors associated with non-response in routine use of patient reported outcome measures after elective surgery in England. *Health Qual Life Outcomes* 2012;10:34

14. Black N, Varagunam M, Hutchings A. Relationship between patient reported experience (PREMs) and patient reported outcomes (PROs) in elective surgery. *BMJ Qual Safety* 2014;23:534-42
15. Neuburger J, Hutchings A, van der Meulen J, Black N. Using patient reported outcomes (PROs) to compare the performance of providers: does the choice of measure matter? *Med Care* 2013;51(6):517-523
16. Black N, Varagunam M, Hutchings A. Influence of surgical rate on patients' reported clinical need and outcomes in English NHS. *J Public Health* 2014;36(3):497-503
17. Varagunam M, Hutchings A, Black N. Relationship between patient reported outcomes of elective surgery and hospital and consultant volume. *Medical Care* (in press)
18. Feng Y, Pistollato M, Charlesworth A, Devlin N, Propper C, Sussex J. Association between market concentration of hospitals and patient health gain following hip replacement surgery. *J Health Sev Res Policy* 2014 doi: 0.1177/1355819614546032
19. Varagunam M, Hutchings A, Black N. Do patient reported outcomes offer a more sensitive method for comparing the outcomes of consultants than mortality? A multi-level analysis of routine data. *BMJ Qual Saf* 10.1136/bmjqs-2014-003551
20. Varagunam M, Hutchings A, Neuburger J, Black N. Impact on hospital performance of introducing routine patient reported outcome measures (PROs) in surgery. *J Health Serv Res Pol* 2014;19:77-84
21. Greenhalgh J, Pawson R, Wright J, Black N, Valderas JM, Meads D, Gibbons E, Wood L, Wood C, Mills C. Functionality and feedback: a protocol for a realist synthesis of the collation, interpretation and utilisation of PROs data to improve patient care. *BMJ Open* 2014;4:e005601 doi:10.1136/bmjopen-2014-005601
22. Van der Wees PJ, Nijhuis-van der Sanden MWG, Ayanian JZ, Black N, Westert GP, Schneider EC. Integrating the use of patient-reported outcomes for both clinical practice and for performance measurement: experts' views from three countries. *Milbank Mem Quarterly* 2014;92:754-75
23. World Health Organization (WHO) (2004). Promoting Mental Health: Concepts, Emerging Evidence, Practice. Geneva: World Health Organization [Online]: http://www.who.int/mental_health/evidence/en/promoting_mhh.pdf (17.01.2013).
24. Kieling, C., Baker-Henningham, H., Belfer, M., Conti, G., Ertem, I., Omigbodun, O., et al. (2011). Child and adolescent mental health worldwide: evidence for action. *The Lancet*, 378(9801), 1515-1525.
25. Wille, N., & Ravens-Sieberer, U. (2010). How to assess resilience: Reflections on a

measurement model. In A. Morgan, E. Ziglio & M. Davies (Eds.), *Health Assets in a Global Context* (pp. 117-144). New York, NY: Springer Verlag.

26. Erhart, M., Ottova, V., Gaspar, T., Jericek, H., Schnohr, C., Alikasifoglu, M., Morgan, A., & Ravens-Sieberer, U. (2009). Measuring mental health and well-being of school-children in 15 European countries using the KIDSCREEN-10 Index. *International Journal of Public Health*, 54 (Suppl. 2), S160-S166. doi: 10.1007/s00038-009-5407-7

27. Mortimer, J., Larson, R. (2002). *The Changing Adolescent Experience: Societal trends and the Transition to Adulthood*, Cambridge University Press, Cambridge.

28. Ravens-Sieberer, U., Wille, N., Erhart, M., Nickel, J., & Richter, M. (2008). Socioeconomic Inequalities in mental health among adolescents in Europe. In World Health Organization (Ed.), *Social cohesion for mental well-being among adolescents* (pp. 26-42). Copenhagen: WHO Regional Office for Europe.

29. Ravens-Sieberer, U., Torsheim, T., Hetland, J., Vollebergh, W., Cavallo, F., Jericek, H., Alikasifoglu, M., Valimaa, R., Ottova, V., & Erhart, M. (2009). Subjective health, symptom load and quality of life of children and adolescents in Europe. *International Journal of Public Health*, 54 (Suppl. 2), S151-S159. doi: 10.1007/s00038-009-5406-8

30. Currie C et al. eds. (2012). *Social determinants of health and well-being among young people. Health Behaviour in School-aged Children (HBSC) study: international report from the 2009/2010 survey*. Copenhagen, WHO Regional Office for Europe (Health Policy for Children and Adolescents, No. 6).

31. Ravens-Sieberer, U., Horka, H., Illyes, A., Rajmil, L., Ottova-Jordan, V., & Erhart, M. (2013). Children's Quality of Life in Europe: National Wealth and Familial Socioeconomic Position Explain Variations in Mental Health and Wellbeing - A Multilevel Analysis in 27 EU countries. *Public Health*, 2013, 1-9. doi: <http://dx.doi.org/10.1155/2013/419530>

32. Barnekow, V., Muijen, M. (2009). Child and adolescent health and development in an European perspective. *International Journal of Public Health*, 54(Suppl. 2), S128-S130.

33. Ottova, V., & Ravens-Sieberer, U. (2010). Social determinants in child health: reflections from the Health Behaviour in School-aged Children survey. *International Journal of Public Health*, 55(6), 525-526. doi: 10.1007/s00038-010-0208-6

34. Maher, I. Waters, E. (2005). Indicators of Positive Mental Health for Children, In: *Promoting Mental Health: Concepts, Emerging Evidence, Practice*, Herrman, Saxena & Moodie, p. 159-168, World Health Organization, 92 4 156294 3, Geneva

35. Ottova, V., Denise, A., Rigby, M., Staines, A., Hjern, A., Leonardi, M. et al. (2013). *Research Inventory of Child Health: A Report on Roadmaps for the Future of Child Health Research in Europe. A European Commission Framework 7 Project 2010 – 2013*. [Online]: <http://www.childhealthresearch.eu/Members/jkilroe/report-on-the-roadmaps-for-the-future-of-european-child-health-research/view> (27.11.2014)

36. Korkeila, J., Tuomi-Nikula, A., Wahlbeck, K., Lehtinen, V., Lavikainen, J. (2006). Proposal for a harmonised set of mental health indicators, In: Improving Mental Health Informatino in Europe, Lavikainen, Fryers & Lehtinen, p. 107-116, Stakes and European Union, Helsinki
37. Office of Disease Prevention and Health Promotion. About Healthy People. Available at: <https://www.healthypeople.gov/2020/About-Healthy-People>. Accessed November 18, 2014.
38. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc*. 2014;21(4):578–582.
39. IOM (Institute of Medicine). 2014. Capturing social and behavioral domains and measures in electronic health records: Phase 2. Washington, DC: The National Academies Press.
40. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Quality of Life Research*. 2009;18(7):873–880.
41. Forrest CB, Margolis P, Seid M, Colletti RB. PEDSnet: How A Prototype Pediatric Learning Health System Is Being Expanded Into A National Network. *Health Aff (Millwood)*. 2014;33(7):1171–1177.
42. Behforouz HL, Drain PK, Rhatigan JJ. Rethinking the social history. *N Engl J Med*. 2014;371(14):1277–1279.
43. PROsetta Stone® Linking Patient Reported Outcome Measures. Available at: <http://www.prosettastone.org>. Accessed October 10, 2015.
44. Valderas JM, Ferrer M, Mendivil J, Rajmil L, Alonso J. Development of a standardized assessment tool for patient reported outcomes: a pilot test with the EMPRO instrument. *Value Health* 2008;11(4):700-8
45. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes: results of the COSMIN study. *Journal of Clinical Epidemiology* 2010;63:737-745.
46. Valderas JM, Alonso J, Prieto L, Espallargues M, Castells FX. 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 13: 35–44
47. 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

48. Forrest CB, Margolis PA, Bailey LC, Marsolo K, Del Beccaro MA, Finkelstein JA, Milov DE, Vieland VJ, Wolf BA, Yu FB, Kahn MG..PEDSnet: a National Pediatric Learning Health System. J Am Med Inform Assoc 2014;21:602- 606. doi:10.1136/amiajnl-2014-002743

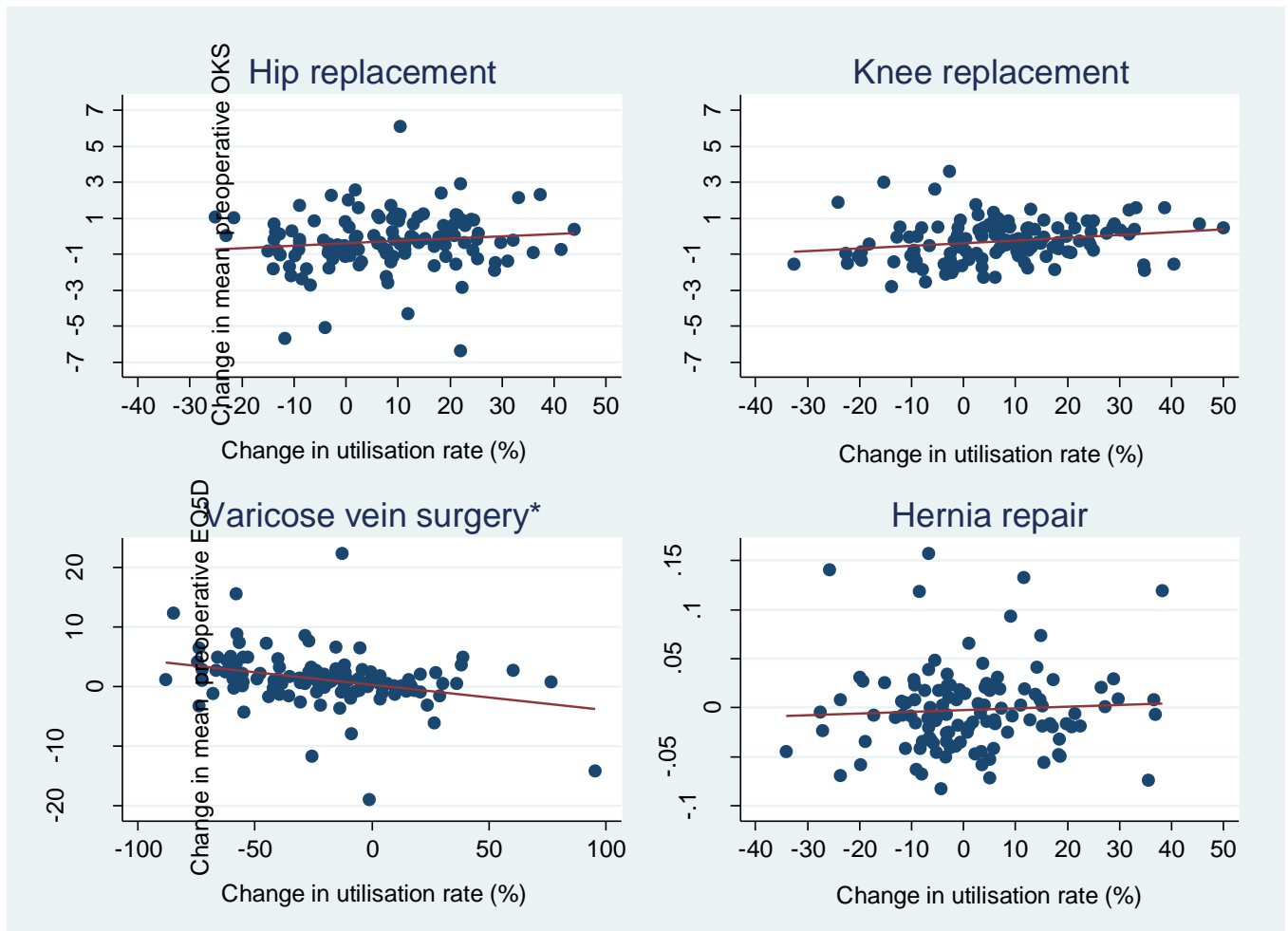
Table 1. U.S. Food and Drug Administration (FDA) roadmap to Patient Focused Outcome Measurement in Clinical Trials.

Understanding the disease or condition	Conceptualizing treatment benefit	Selecting/developing the outcome measure
<p>Natural history of the disease or condition</p> <ul style="list-style-type: none"> • <i>Onset, duration, resolution</i> • <i>Diagnosis</i> • <i>Pathophysiology</i> • <i>Range of manifestations</i> 	<p>Identify concept(s) of interest for meaningful treatment benefit, i.e., how a patient:</p> <ul style="list-style-type: none"> • <i>Survives</i> • <i>Feels (e.g., symptoms)</i> • <i>Functions</i> 	<p>Search for existing clinical outcome assessment measuring concept(s) of interest in context of use</p> <ul style="list-style-type: none"> • <i>Measure exists</i> • <i>Measure exists but needs to be modified</i> • <i>No measure exists</i> • <i>Measure under development</i>
<p>Patient subpopulations</p> <ul style="list-style-type: none"> • <i>By severity</i> • <i>By onset</i> • <i>By comorbidities</i> • <i>By phenotype</i> 	<p>Define context of use for clinical trial</p> <ul style="list-style-type: none"> • <i>Disease/Condition entry criteria</i> • <i>Clinical trial design</i> • <i>Endpoint positioning</i> 	<p>Begin clinical outcome assessment development</p> <ul style="list-style-type: none"> • <i>Document content validity (qualitative or mixed methods research)</i> • <i>Evaluate cross-sectional measurement properties (reliability and construct validity)</i> • <i>Create user manual</i> • <i>Consider submitting to FDA for clinical outcome assessment qualification for use in exploratory studies</i>
<p>Health care environment</p> <ul style="list-style-type: none"> • <i>Treatment alternatives</i> • <i>Clinical care standards</i> • <i>Health care system perspective</i> 		
<p>Patient/caregiver perspectives</p> <ul style="list-style-type: none"> • <i>Definition of treatment benefit</i> • <i>Benefit-risk tradeoffs</i> • <i>Impact of disease</i> 	<p>Select clinical outcome assessment type</p> <ul style="list-style-type: none"> • <i>Patient-Reported Outcome (PRO)</i> • <i>Observer-Reported Outcome (ObsRO)</i> • <i>Clinician-Reported Outcome (ClinRO)</i> • <i>Performance Outcome (motor, sensory, cognition)</i> 	<p>Complete clinical outcome assessment development:</p> <ul style="list-style-type: none"> • <i>Document longitudinal measurement properties (construct validity, ability to detect change)</i> • <i>Document guidelines for interpretation of treatment benefit and relationship to claim</i> • <i>Update user manual</i> • <i>Submit to FDA for clinical outcome assessment qualification as effectiveness endpoint to support claims</i>

Table 2. Common metrics, and measurement tools for selected conceptual domains

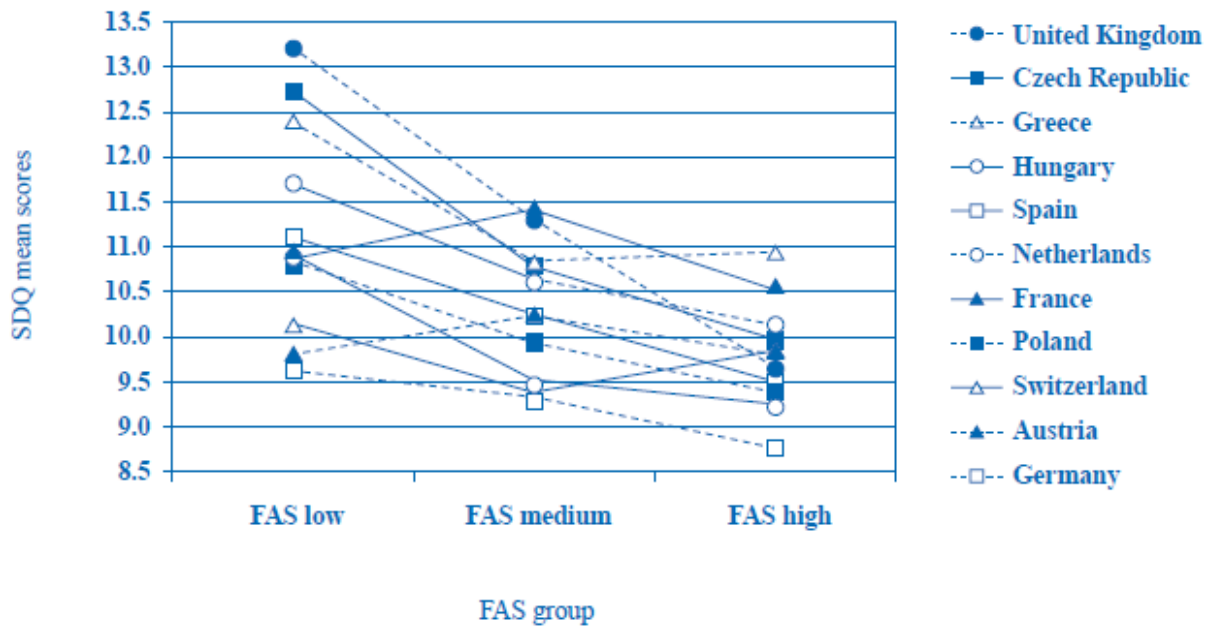
Conceptual Domain	Common Metric	Measurement Tools
Length	Meters	<ul style="list-style-type: none"> • Metric ruler • Imperial unit ruler
Blood Pressure	Millimeters of mercury	<ul style="list-style-type: none"> • Manual sphygmomanometers • Automated sphygmomanometers • Automated devices for self-measurement at the wrist
Physical Activity	Metabolic Equivalent of Task (MET) minutes	<ul style="list-style-type: none"> • Self-reported questionnaires • Actigraphs • Accelerometers
Depressive Symptoms	PROMIS T-Score	<ul style="list-style-type: none"> • PROMIS Depression Short Form • Patient Health Question-9 (PHQ-9) • Beck Depression Inventory-II

Figure 1. Association between change between 2009/10 and 2011/12 in district utilisation rate and change in pre-operative severity (mean score in Patient Reported Outcome scores).



OHS: Oxford Hip Score; OKS: Oxford Knee Score; AVVQ: Aberdeen Varicose Veins Questionnaire; EQ5D: EurQoL 5D.

Figure 2. Mental health problems and family affluence.



FAS: Family Affluence Scale; SDQ: Strengths and Difficulties Questionnaire. Reproduced with permission from Ravens-Sieberer et al. 2008, p.37