Review

Patient Report Outcome Measures in Orthopaedics[1]

Joel J. Gagnier ND, MSc, PhD[1-2]

1. Department of Orthopaedic Surgery, University of Michigan Health System, Ann Arbor, MI
2. Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI

Correspondence: Please direct all correspondence to Dr. Joel J. Gagnier, email: jgagnier@umich.edu.

[1] This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi:10.1002/jor.23604

Abstract

Patient reported outcome measures (PROMs) are key tools when performing clinical research and PROM data are increasingly used to inform clinical decision-making, patient-centered care, health policy and more recently, reimbursement decisions. PROMs must possess particular properties before they are used. Thus purpose of this paper is to give an overview of PROMs, their definition, how their evidence can be assessed, how they should be reported in clinical research, how to choose PROMs, the types of PROMs available in orthopaedics, where these measures can be found, PROMs in orthopaedic clinical practice and what are some key next steps in this field. If PROMs are used in accordance with the guidance in this article, I believe we will gain considerable insight into PROMs in orthopaedics and will advance this field in a way that can contribute to science, improve patient care and save considerable resources.

Why Patient Reported Outcome Measures?

The development, testing and implementation of tools to aid in the measurement of phenomena in medicine are central to clinical practice and clinical research. Measurements in clinical practice form the basis of diagnosis, prognosis, evaluation and follow-up. Measurements in clinical research allow for the collection of data that afford us the information needed to test specific hypotheses [1]. The field of measurement in medicine includes both psychometrics and clinimetrics [2-4]. But, it has been argued that there is little distinction between these two areas [3]. Throughout this paper the term psychometrics will be used and more generally the term measurement to refer to these fields.

There are many types of measurements used in medicine including subjective (e.g., patient, clinician reported) and objective (e.g., imaging or laboratory tests) types of measures. Patient reported outcome measures (PROMs), also called self-report measures, propose to collect information related to constructs that are reported by the patients themselves, without interpretation by other parties [1]. Patient reported outcomes include perceptions and opinions on symptoms, functioning, health–related quality of life (HRQoL), and satisfaction, among other areas (Figure 1) [1,5]. PROMs different focuses including: generic health-related (e.g., Medical Outcomes Survey Short Form-36), disease/diagnosis specific (e.g., Western Ontario Rotator Cuff Index), or regionally specific e.g., American Shoulder and Elbow Society score).

Insert Figure 1 Here

Patient reported outcome measures (PROMs) are key tools when performing clinical research and PROM data are increasingly used to inform clinical decision-making, patient-centered care, health policy and more recently, reimbursement decisions [6-8]. Furthermore, substantial federal fiscal commitment has been directed at developing PROMs [e.g., Patient Reported Outcome Measurement Information System (PROMIS); 9-17]. Evidence suggests that the use of PROMs has increased dramatically over the last 20 years and it has been noted that this trend will continue [18]. Therefore, the use of PROMs must be tempered and directed by solid evidence and appropriate guidance.

insert figure 2 here

Quality Assessment of Patient Reported Outcome Measures

Health status measurement instruments must possess adequate measurement properties (e.g., reliability, validity, responsiveness) to be useful for research or patient

care [2; 21-23]. A measurement tool is considered to be valid when it actually measures what it proposes to measure. There are several types of validity including: content, construct, and criterion validity. Reliability is the property of measuring some phenomenon in a predictable manner (repeatability). Some forms of reliability include: internal consistency, test-retest or intra-rater, and inter-rater reliability. Finally, responsiveness is the ability to detect change in the underlying construct over time, even if the changes are small [21-23]. Table 1 describes definitions of all measurement properties.

insert table 1 here

Outcomes used in patient centered outcomes research, such as PROMs, should be of high quality so as to ensure that these measures do not introduce bias in the effect estimates for these outcomes. For example, poor quality PROMs used in effectiveness studies will lead to unreliable and misleading results of these studies, potentially resulting in harm to the patient or the inappropriate use of resources (e.g., health care dollars, research funding allocation). PROMs must have appropriate measurement properties [see table 1] to detect small but relevant treatment effects or changes in individual and groups of patients. Only then can we trust the results of research that includes such PROMs to inform decision-making.

To properly evaluate the quality of psychometric properties of a PROM one should first search for or conduct a systematic review of existing evidence, and then secondarily initiate new primary research to focus on areas where the measure is either flawed or where no evidence is available. Searching for PROMs involves a careful use of available electronic databases including general databases such as MEDLINE, Cochrane

Library, and EMBASE as well as topic specific databases such as SportDiscus and online warehouses of such systematic reviews such as the COSMIN database of systematic reviews of measurement instruments (http://database.cosmin.nl/). For conducting new systematic reviews of evidence for specific PROMs, there have been several methods proposed for evaluating the properties of a PROM [24-26].

The most widely used and recent set of criteria for assessing the measurement properties of PROMs is the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) which include detailed sets of items for assessing the methodological quality of research that tests a psychometric quality [25]. The COSMIN checklist includes sets of items for assessing a study's methodological quality that proposes to assess one of 9 measurement properties (i.e., internal consistency, reliability, measurement error, content validity, structural validity, hypothesis testing, cross-cultural validity, criterion validity and responsiveness) as well as two additional boxes to extract information regarding interpretability and generalizability, which strictly are not called measurement properties, and finally one box to assess item response theory (IRT) methods if used [25-26]. Each of the 9 measurement properties include a selection of items that are scored on a 4 point Likert scale of methodological quality as excellent, good, fair, poor quality [27-28]. The COSMIN checklist is the most comprehensive checklist available, was developed using strong methods and is widely used [25-26].

When performing systematic reviews of the psychometric evidence of PROMs one must be careful to assess both the methodological quality of each the primary studies and also assess the evidence for or against the psychometric property itself [1,29-31]. For example, if a published study investigates the inter-rater reliability of an instrument, that

study must have appropriate methods of doing so (e.g., a proper sample size calculation for the reliability statistic) and also the level of the reliability statistic must meet some cut-off (e.g, intraclass correlation coefficient / weighted Kappa ≥ 0.70 OR Pearson's r ≥ 0.80). A comprehensive set of criteria for assessing the psychometric evidence of health status measurement instruments were proposed by Terwee et al (2007)[22](see table 1)[29,32]. Each criterion is rated as positive (+), indeterminate (?; unclear from what is reported), or negative (-). If no information for a property is available in the literature, a rating of zero (0) would be given to indicate no evidence.

Finally, the methodological quality of the included studies and the psychometric evidence should be synthesized to arrive at an overall rating. Schellingerhout et al (2012)[32] proposed a synthesis method which combines the consistency of the psychometric evidence with the methodological quality of the included studies and the level of evidence proposed by the Cochrane Back Review Group [33]. The overall results are then categorized as positive (+), unknown/indeterminate (?), negative (-), or no evidence (0) accompanied with the overall level of evidence ranging from unknown to strong (see Table 2). A rating of conflicting results (-/+) is given when the number of positive ratings equals the number of negative ratings. Using this method, when combined across studies, the levels of evidence are: strong (representing consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality), moderate (representing consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality), limited (representing one study of fair methodological quality), conflicting (representing

conflicting findings) and unknown (representing the existence of only studies of poor methodological quality).

insert table 2 here

Choosing a PROM

Thus when attempting to choose a PROM for use clinically or research in orthopaedics, and any other area, one must be careful to identify an appropriate instrument for their specific purposes (e.g, condition/disease/diagnosis specific PROM, regionally specific PROM, generic health/QOL PROM) and then must collect evidence for the quality of that instrument. Just because an instrument is used frequently in the published literature does not guarantee its quality. In fact, there is extensive evidence, which we review below, that PROMs used in many areas of orthopaedics are significantly flawed, or at least, lack data on their psychometric properties [e.g., 29-31]. The use of poor quality instruments will result in biased or unreliable effect estimates and can potentially mislead decision makers relying this evidence as well as harm patients and waste resources.

There is some empirical evidence that clinical trials which use poor or unknown quality PROMs are more likely to report that a treatment was superior to control, by up to 89% (RR 1.89, 95% CI 1.40-2.56), and the effect was even greater in non-pharmacologic trials, 168% (RR 2.68, 95% CI 1.86 – 3.84) [e.g., 34]. Furthermore, clinical trials that include PROMs frequently do not report any information on the measure's psychometric properties [e.g., 35].

The issue of underreporting is a problem in many sections of published clinical trials and has spurred on much literature in the form of reporting guidelines and

recommendations aimed at improving the completeness of published scientific reports and protocols (http://www.equator-network.org/library/). In particular, an extension of the Consolidated Standards of Reporting Trials (CONSORT) for patient reported outcomes (PRO) was developed to guide the reporting of PROs in randomized clinical trials (RCTs; CONSORT-PRO) [36]. The CONSORT-PRO checklist contains 5 items to refer to when reporting the use of a PRO as a primary and secondary outcome measure. Related to this effort, a task force of the International Society of Quality of Life (ISOQOL) Research created a suite of reporting standards for HRQOL outcomes in RCTs [37]. Furthermore, ISOQOL developed a set of minimum standards for the design and selection of PROMs for use in patient-centered outcomes research (PCOR) and comparative effectiveness research (CER) [38]. These can be used in conjunction with COSMIN criteria when choosing a PROM for use in research. In some cases, the evidence for a PROMs properties may not exist, but this should not prevent us from doing research. In addition, guidance for the reporting of PROs in clinical trial protocols is currently under construction [39]. These standards should be adhered to when choosing a PRO for clinical research and when reporting that research in the form of peer-reviewed manuscripts.

PROMs in Clinical Research

The process of selecting a PROM for patient monitoring or clinical research must be considered carefully. There is empirical evidence that choosing an instrument on the basis of frequency of use in the literature does not guarantee its quality. For example, in a systematic review of PROMs used in clinical publications of patient's following knee arthroplasty, the Knee Society Score (KSS) and the Western Ontario McMaster

Osteoarthritis Index (WOMAC) were used in over 80% of the 438 included articles [40]. But a recently published rigorous systematic review of PROMs developed and tested in patients undergoing or who have undergone total knee arthroplasty found that the KSS and WOMAC exhibited less than favorable psychometric properties [30]. This suggests there is a large volume of published clinical research that use non-valid, unreliable and unresponsive instruments to measure patient reported outcomes, and therefore any such findings are not trustworthy. Of course, this is just one example (i.e., knee arthropalsty) from the literature and additional rigorous work exploring this hypothesis is needed [41].

With that caution in mind, when choosing a PROM for clinical research or monitoring of patients in clinical practice you need to determine what domains or constructs you want to measure (e.g., pain, physical function, quality of life), what types of patients you plan to include (e.g., patients who have undergone total hip replacements or have a specific diagnosis), if an outcome measure exists that captures these domains for these patients, what the psychometric evidence is for the measure or measures you have identified, and finally if the test is usable (i.e., what is the test burden, time it takes to complete and are the scores sensible and interpretable). This is not a trivial process obviously as each of these steps will take time and concerted effort. The ISOQOL standards (see table 3) should be referred to when choosing a PROM for research and can also be used as a guide for the choice of PROMs in clinical practice [39]. Furthermore, below we describe how to assess and use systematic reviews of the psychometric properties of instruments.

insert table 3 here

Core Outcome Measurement Sets

Furthermore, one should also consider looking for a core outcome set (COS) in the area of research they planning. A COS is defined as an agreed upon minimum selection of outcomes that should be measured and reported in all clinical trials for a particular health condition [42]. There is an expectation that the core set of outcomes would always be collected and reported, but it would not preclude use of additional outcomes in particular clinical research. It is argued that the lack of uniformity in outcome measurement across trials limits our ability to compare findings between studies or to pool data for meta-analyses. Also, selective outcome reporting (i.e., selective reporting of favorable or statistically significant outcomes) can bias the results of systematic reviews [43]. Thus, in an effort to reduce heterogeneity in outcomes measured across clinical trials, the development of core COS in specific health conditions has been recommended [44]. A COS would increase the use of important outcomes (i.e., important to all relevant stakeholders; e.g., patients, clinicians, researchers), outcomes with sufficient psychometric evidence for their use, improve the validity of outcome data for these outcomes in clinical research and increase the feasibility of conducting meta-analyses on such topics [42,45].

The Core Outcome Measures in Effectiveness Trials (COMET) [42] and the Outcome Measures in Rheumatology (OMERACT) [46] initiatives provide methodological guidance, including a stepwise approach, for the development of a COS [46,47-48]. Several examples of COSs that have been developed for orthopaedic conditions and procedures include: degenerative lumbar conditions [49], non-specific low back pain [50], post-surgical knee pain [51], hip fractures [52], shoulder pain [53-54], and distal radius fractures [55]. There are many additional areas where COSs are needed

in orthopaedics. Of course, a frequent component of a COS is a PROM, thus when planning on conducting a clinical trial and seeking a PROM it is wise to search for a COS in your area of inquiry. The COMET group houses a database of COSs on their website.

Furthermore, beyond disease (e.g., rotator cuff tear) or regionally specific (e.g., shoulder) PROMs one might consider other more general PROMs such as those that measure overall HRQOL (e.g., EuroQol EQ-5D) since these measures may give some determination of the overall impact of the condition of the person's quality of life. This might allow a comparison between conditions as to their overall impact on individuals [56].

There are many instruments available, many PROMs, that can be used in most areas of orthopaedics [e.g., 29-31,49-55]. But, if an appropriate PROM does note exist, after an exhaustive literature search for said instrument, then one may embark on developing a new instrument specific to their needs. When developing a new instrument it is imperative that proper methods be followed. It is not within the scope of this paper to describe these, but many excellent textbooks are available to help those that are interested [e.g., 1-2, 5, 57-58].

Systematic Reviews of Psychometric Evidence for PROMs in Orthopaedics

There has been some progress made in systematically evaluating the psychometric properties of a patient reported outcome measure for use in orthopaedics. Here I list some of the systematic reviews in this area that use relatively accepted methods. But this is not a comprehensive systematic look at all the evidence of PROMs in orthopaedics. As noted above, searching for PROMs involves a careful use of available electronic databases including general databases such as MEDLINE, Cochrane Library, and EMBASE as well

as topic specific databases such as SportDiscus and online warehouses of such systematic reviews such as the COSMIN database of systematic reviews of measurement instruments (http://database.cosmin.nl/).

While this is not a comprehensive list, systematic reviews of the psychometric evidence for PROMs for use in orthopaedic related conditions have been completed for the knee [59], total knee replacement [30], anterior cruciate ligament injuries [60], total hip replacement [61], foot and ankle disorders [31,62], chronic ankle instability [63], rotator cuff disease [29], shoulder disability [64,65], shoulder specific PROMs [66], shoulder functional scores [67], functional limitations in the athletic shoulder [68], shoulder disability measures translated in Portuguese [69], brachial plexus injuries [70], upper extremity following trauma [71], elbow [72], hand injuries [73, 74], carpal tunnel syndrome [75], wrist and hand function [76], wrist [77], neck pain [78, 79], non-specific neck pain [80], cervical degenerative disease [81], cervical pain or dysfunction [82], neck-specific questionnaires in other languages [83], non-specific low back pain [84,85]. But many more such systematic reviews are available for the array of conditions that are seen by orthopaedic surgeons.

When reading such systematic reviews, be careful to be critical of the methods used in them to evaluate the instruments. A systematic review of a sample of 148 reviews found that a very small proportion of them, 7.4%, used proper methods for evaluating the psychometric evidence [86]. While this review is a bit dated (2009), and given that the COSMIN and Evaluating the Measurement of Patient Reported Outcomes (EMPRO) guidance were published after this, I would expect that more recent reviews of psychometric properties of PROMs have better quality, but we recommend that this

hypothesis be explored. In the end, make sure to use some explicit criteria when assessing systematic reviews of the psychometric evidence for PROMs. Effort should be made to rigorously develop critical appraisal criteria for systematic reviews of measurement properties of PROMs. I provide some preliminary guidance in table 4.

insert table 4 here

PROMs in Orthopaedic Clinical Practice

The use of PROMs across large samples can lead to broad sweeping quality improvement initiatives, but the use of PROMs on a patient by patient basis, to inform clinical decision making during patient follow-up, has its unique challenges. PROMs have been used in clinical practice for many decades, whether it be a pain scale or some more elaborate measure such as a depression index. In orthopaedic surgery there is an array of PROMs that could potentially be used to manage individual patient care, some of which are cited above [39-31, 60-85], but many more exist and may be appropriate for a specific clinical focus. The two most comprehensive databases of individual PROMs are PROQOLID [87] and BIBLIOPRO [88], the latter of which is a database of instruments in Spanish. Also, there is an array of instruments offered through the PROMIS initiative, which I describe in more detail here.

PROMIS is an NIH Common Fund project involving the dynamic assessment of PROs. PROMIS includes item banks that measure key health symptoms/concepts for both the general population and several chronic conditions. PROMIS item banks assess physical (physical function, fatigue, sleep disturbance, sleep related impairment, pain behavior, and pain interference), emotional (depression, anxiety, and anger), cognitive (applied cognition-abilities and applied cognition general concerns), and social health

(ability to participate, satisfaction with social roles and activities, emotional support, instrumental support, informational support, and social isolation; see Appendix A for a list of item banks and sample items). These item banks were developed following rigorous protocols that involved extensive formative research and statistical analysis [9-14]; state-of the art psychometric analysis was used to develop these measures (including classical test theory approaches and item response theory; IRT). Items can be administered as a full set, a computerized adaptive test (CAT), or a calibrated "short form" (preselected set of items). An item bank that can be administered as a CAT provides an advantage to any preselected set of items (i.e., short form) in maximizing assessment sensitivity while simultaneously minimizing the number of items. PROMIS is available, free of charge, through www.assessmentcenter.net. PROMIS health assessment data can be collected with minimal respondent burden and without error introduced by data entry through the use of electronic systems [89]. While more research needs to be done on the measurement properties of PROMIS instruments in orthopaedic samples, if a clinician can manage to set up electronic data collection then using a PROMIS instrument is certainly recommended.

How PROMs contribute to Clinical Monitoring

There is of course a great opportunity for the use of PROMs to aid in the monitoring and care of individual patients. For example, PROMs can aid in the screening, diagnosis, prognostic monitoring, and assessment of treatment interventions while promoting patient-centered care [90]. But, as Porter et al (2016) point out in their excellent paper on the framework and guidance for implementing patient-reported outcome in clinical practice, there are many more potential clinical applications [90].

Some of these include: Supporting decision-making in the diagnostic process (e.g., Screening, Diagnosis), informing risk stratification and prognosis (identification of vulnerable patients and patients 'at risk'), supporting prioritization and goal setting, supporting decision-making in indication for treatment (medical/surgical), facilitating monitoring of general health status, health status related to the specific diagnosis/risk, response to treatment/management, facilitating communication between patients and health professionals and within teams and between professionals (i.e., consistent use of PROMs along the entire care pathway) [90].

Challenges of using PROMs in Clinical Practice

There are many challenges confronting the routine use of PROMs in clinical practice [90,91]. For example, clinicians might view the addition of PROMs to clinical practice disruptive or burdensome and clinicians may not envision how to use PROMs in their practice or may lack the expertise to interpret and apply the scores of information derived from such measures [92-94]. Therefore, training programs for clinicians in these areas are needed. Furthermore, many question the efficacy or impact that feedback from PROMs can have patient care and outcomes. In fact, the evidence for the impact of using PROMs in clinical practice is scarce. While several systematic reviews have explored this area [95-108], most have been reported to be inconclusive and this is mostly due to methodologic flaws of the primary studies [90]. That is, more research is needed, especially in orthopaedics where there are virtually no rigorous studies exploring these questions.

Where to go from here?

Then where does all of this leave us in relation to how we might find, use and improve PROMs in clinical research and practice? I have chosen to frame this last section in the form of what I view are the most important needs facing the use of PROM in clinical research and practice in orthpaedics: 1. The systematic organization of the field of measurement in orthopaedics to highlight the best PROMs, this would involve performing rigorous systematic reviews of the existing evidence and new psychometric research on existing instruments. Also, there should be no new development of PROMs unless an obvious gap has been identified through a systematic review. 2. The creation of core outcome measurement sets in all areas of orthopaedics to help guide clinical researchers. The creation of core outcome sets must proceed through a rigorous approach and be endorsed by relevant research groups (e.g., OMERACT, COMET) and clinical organizations (e.g, AOSSM etc). 3. Rigorous research to determine how PROMs are, or are not, useful to clinical management. For example, we require much research to be be done to determine how clinicians can use PROMs when monitoring patients, and what are the most efficient and effective methods to implement them in clinical practice. Also, empirical evidence must be generated on how the use of PROMs may help with care decisions and how to present PROM data in electronic charts., 4. Increased funding for research in all of these areas. It is obvious that the disorganization of this area of inquiry in orthopaedics obviates the need for systematic and rigorous research that should be supported federally and through other funding bodies (e.g., foundations). Some funding should go to supporting individual research projects but also to supporting researchers themselves to gain new experience in this area. 5. Training programs for those interested in advancing research in PROMs and for clinicians seeking to understand their use and

training at the resident or graduate student level. Formal training programs and course will support clinicians and researchers in gaining key knowledge to allow them to further implement and refine PROMs and the related methodologic research that is needed.

Conclusions

PROMs have an obvious place in othropaedics in that they reflect underlying constructs of direct importance to and as reported by patients themselves. These measures are a key component to clinical research in orthopaedics and may also be so for clinical practice in orthopaedic surgery. If some of the needs can be addressed as expressed above, I believe we will gain considerable insight into PROMs in orthopaedics and will advance this field in a way that can contribute to science, improve patient care and save considerable resources.

References

1. deVet HCW, Terwee CB, Mokkink LB, Knol DL. 2011. Measurement in Medicine. Cambridge: University printing House; 337 p.
2. Streiner DL, Norman GR, Cairney J. 2008. Health Measurement Scales: a practical guide to their development and use, 5th ed. Oxford: Oxford University Press; 399 p.
3. Streiner Dl. Clinimetrics vs. psychometrics: an unnecessary distinction. J Clin Epidemiol. 2003;56:1142-5.
4. Feinstein AR. Clinimetrics. New Haven, CT: Yale University Press; 1987.
5. Fayers PM, Machin D. 2007. Quality of Life, 2nd ed. Wiley & Sons: Chichester; 544 p.

6. FDA Guidance on Patient Reported Outcomes (PROs) http://www.ispor.org/workpaper/FDA%20PRO%20Guidance.pdf. Accessed Oct 15,2015.
7. Basch E. The missing voice of patients in drug safety reporting. N Engl J Med. 2010;362(10): 865-869.
8. Lipscomb J, Reeve BB, Clauser SB, et al. Patient reported outcomes assessment in cancer trials: taking stock, moving forward. J Clin Oncol. 2007; 25(32):5133-5140.
9. DeWalt, D. A., Rothrock, N., Yount, S., & Stone, A. A. Evaluation of item candidates: The PROMIS Qualitative Item Review. Medical Care, 45(5), S12–S21.
10. Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Terisi, J. A., Thissen, D., Reviki, D. A., Weiss, D. J., Hambleton, R. K., Liu, H., Gershon. R., Reise, S. P., Lai, J. S., & Cella, D. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the patient-reported outcomes measurement Information system (PROMIS). Medical Care, 45(5), S22–S31.
11. Liu HH, Cella D, Gershon G, Ahen J, Morales LS, Riley W & Hays RD (2010). Representativeness of the PROMIS internet panel. Journal of Clinical Epidemiology, 63(11), 1169-1178.
12. Rothrock NE, Hays RD, Spritzer K, Yount SE, Riley W, & Cella D (2010). Relative to the general US population, chronic diseases are associated with poorer health-related quality of life as measured by the Patient-Reported Outcomes Measurement Information System (PROMIS). Journal of Clinical Epidemiology, 63(11), 1195-1204.

13. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, Amtmann D, Bode R, Buysse DJ, Choi SW, Cook KF, DeVellis R, DeWalt D, Fries JF, Gershon R, Hahn E, Pilkonis P, Revicki D, Rose M, Weinfurt K & Hays R (2010). The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. Journal of Clinical Epidemiology, 63(11), 1179-1194.

14. PROMIS standards http://www.nihpromis.org/Documents/PROMIS_Standards_050212.pdf

15. Nahm ML, Pieper CF, Cunningham MM (2008) Quantifying data quality for clinical trials using electronic data capture. PLoS ONE 3(8): e3049. doi:10.1371/journal.pone.0003049

16. Gershon RC, Rothrock N, Hanrahan R, Bass M, Cella D. (2010). The use of PROMIS and Assessment Center to deliver patient-reported outcomes in clinical research. Journal of Applied Measurement, 11(3): 304-314.

17. Gershon RC, Rothrock NE, Hanrahan RT, Jansky LJ, Harniss M, Riley W. (2010). The development of a clinical outcomes survey research application: Assessment CenterSM. Quality of Life Research. 19(5):677-685.

18. Sanders C, Egger M, Donovan J, Tallon D, Frankel S: Reporting on quality of life in randomised controlled trials: bibliographic study. BMJ 1998, 317:1191–1194.

19. Soreide K, Soreide AH. Using patient reported outcome measures for improved decision making in patients with gastrointestinal cancer- the last clinical frontier in surgical oncology? Frontiers in Oncology. 2013;3:157.

20. Bryan S, Whitehurst D. Patient-reported outcome measurement (PROMs) in British Columbia: What has been achieved so far and where next? Quality Forum 2013, Vancouver Coastal Health Research Institute.
21. DeVellis RF. Scale development: Theory and applications (3rd Ed). 2012.
22. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60:34-42.
23. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010 Jul;63(7):737-45.
24. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, Stein RE. Assessing health status and quality of life instruments: attributes and review criteria. Qual Life Res. 2002;11(3):193-205.
25. Valderos JM, Ferrer M, Mendivil J, et al. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. Value Health 2008;11:700-8.
26. Mokkink LB et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19:539-549.
27. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21(4):651.

28. COSMIN Checklist with 4-point scale. http://www.cosmin.nl/cosmin_checklist.html Accessed January 4th, 2017.

29. Huang H, Grant JA, Miller BS, et al. A Systematic Review of the Psychometric Properties of Patient-Reported Outcome Instruments for Use in Patients With Rotator Cuff Disease. Am J Sports Med 2015;43(10):2572.

30. Gagnier JJ, Mullins M, Huang W. Patient Reported outcome questionnaires for total knee arthroplasty: A systematic review of psychometric properties and methodological quality. J of Arthroplasty. 2016 (Accepted).

31. Jia Y, Huang H, Gagnier JJ. Patient Reported outcome questionnaires for foot and ankle conditions: A systematic review of psychometric properties and methodological quality. Qual of Life Res. 2016 (Accepted).

32. Schellingerhout JM, Verhagen AP, Heymans MW, et al. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. Qual Life Res 2012;21(4):659

33. van Tulder M, Furlan A, Bombardier C, et al. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. Spine (Phila Pa 1976) 2003;28(12):1290.

34. Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales: a major source of bias in randomized controlled trials of treatments for schizophrenia. Br J Psychiatry. 2000;176:249-252.

35. Mercieca-Bebber RL, Perreca A, King M et al. Patient-reported outcome in head and neck and thyroid cancer randomized controlled trials: A systematic review of

completeness of reporting and impact on interpretation. Europ J Cancer. 2016;56:144-161.

36. Calvert M, Blazeby J, Altman DG et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. JAMA 2013; 309: 814–822.
37. Brundage M, Blazeby J, Revicki D et al. Patient-reported outcomes in randomized clinical trials: development of ISOQOL reporting standards. Qual Life Res. 2013;22:1161-1175.
38. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, Butt Z. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res. 2013, 22(8), 1889-1905.
39. SPIRIT-PRO Extension. http://www.equator-network.org/2016/10/10/spirit-pro-extension-stakeholder-survey-invitation/. Accessed January 6th, 2017.
40. Theodoulou A, Bramwell DC, Spiteri AC et al. The use of scoring systems in knee arthroplasty: A systematic review of the literature. J Arthroplasty. 2016;21:2364-2370.
41. Gagnier JJ, Robbins C, Huang W. Patient reported outcome measures for rotator cuff disease: quality and potential for biased estimates of treatment effect. (personal communication).
42. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, Tugwell P. Developing core outcome sets for clinical trials: issues to consider. Trials. 2012; 13:132.

43. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, Williamson PR. The impact of outcome reporting bias in randomized controlled trials on a cohort of systematic reviews. BMJ. 2010;340:c365.

44. Clarke M: Standardising outcomes for clinical trials and systematic reviews. Trials 2007;8:39.

45. Kirkham JJ, Gargon E, Clarke M, Williamson PR. Can a core outcome set improve the quality of systematic reviews? A survey of the Coordinating Editors of Cochrane Review Groups. Trials. 2013;14:21.

46. Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: an international initiative to improve outcome measurement in rheumatology. Trials. 2007;8:38.

47. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, D'Agostino MA, Conaghan PG, Bingham CO, Brooks P, Landewé R, March L, Simon LS, Singh JA, Strand V, Tugwell P. Developing core outcome measurement sets for clinical trials: OMERACT Filter 2.0. J Clin Epidemiol. 2014;67:745–753.

48. Boers M, Kirwan JR, Gossec L, Conaghan PG, D'Agostino MA, Bingman CO 3rd, Brooks PM, Landewe R, March L, Simon L, Singh L, Singh JA, Strand V, Wells GA, Tugwell P. How to choose core outcome measurement sets for clinical trials: OMERACT 11 approves Filter 2.0. J Rhuematol. 2014;41(5):1025-30.

49. Clement RC, Welander A, Stowell C et al. A proposed set of metrics for standardized outcome reporting in the management of low back pain. Acta Orthopaedica. 2015;86:523-533.

50. Chiarotta A, Deyo RA, Terwee CB et al. Core outcome domain for clinical trials in non-specific low back pain. Europ Spine J. 2015;24(6):1127-1142.

51. Wylde V, Mackichan F, Bruce J, Gobberman-Hill R. Assessment of chronic post-surgical pain after knee replacement: Development of a core outcome set. Europ J Pain. 2014;19(5):611-620.

52. Haywood KL, Griffin XL, Achten J, Costa ML. Developing a core outcome set for hip fracture trials. Bone Joint J. 2014;96-B(8);1016-1023.

53. Buchbinder R, Page M, Huang H, Verhagen A, Beaton D, Kopkow C, Lenza M, Nitin B Jain NB, Richards B, Richards P, Voshaar M, van der Windt D, Gagnier JJ, for the Shoulder Core Outcome Set Special Interest Group. A preliminary core domain set for clinical trials of shoulder disorders: A report from the OMERACT 2016 Shoulder Core Outcome Set Special Interest Group. J of Rheumatology. 2017 (Accepted).

54. Page M, Huang H, Verhagen A, Buchbinder R, Gagnier JJ. Domains for a core outcome set for clinical trials of shoulder pain: Results of an international Delphi process. RMD Open. 2017 (Accepted).

55. Doldhahn JB, Ladd D, Macdermid A et al. Recommendation for measuring clinical outcome in distal radius fractures: a core set of domains for standardized reporting in clinical practice and research. Arch Orthopaed Trauma Surg. 2014;134(2);197-205.

56. Black N. Patient reported outcome measures could help transform healthcare. BMJ. 2013;346:19–21.

57. DeVeliis RF. 2012. Scale Development: theory and applications. 3rd ed. Los Angeles: Sage; 205 p.

58. Cappelleri JC, Zou KH, Bushmakin AG et al. 2014. Patient-reported outcomes: measurement, implementation and interpretation. Boca Raton: Taylor & Francis Group; 331 p.

59. Wang D, Jones MH, Khair MM, Miniaci A. Patient-reported outcome measures for the knee. J Knee Surg. 2010;23(3):137151.

60. Shen Y, Qiang W, Gagnier JJ. Patient reported outcome measures for anterior cruciate ligament injuries: A systematic review of psychometric evidence. 2016 (submitted).

61. Gagnier JJ, Huang W, Mullins M. Patient Reported outcome questionnaires for total hip arthroplasty: A systematic review of psychometric properties and methodological quality. 2016 (submitted).

62. Van der Leeden M, Steultjens MP, Terwee CB et al. A systematic review of instruments measuring foot function, foot pain, and foot-related disability in patients with rheumatoid arthritis. Arthritis Rheum. 2008;15(9):1257-1269.

63. Eechaute C, Vaes P, Van Aerschot L, Asman S, Duquet W. The clinimetric qualities of patient-assessed instrument for measuring chronic ankle instability: A systematic review. BMC Musculoskelet Disord. 2007;8:6.

64. Michener LA, Leggin BG. A Review of self-report scales for the assessment of functional limitation and disability of the shoulder. J Hand Ther. 2001;14(2):68-76.

65. Bot SD, Terwee CB, van der Windt DA et al. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. Ann Rhuem Dis. 2004;63(4):335-341,

66. Schmidt S, Ferrer M, Gonzalez M et al. Evaluation of shoulder-specific patient-reported outcome measures: a systematic and standardized comparison of available evidence. J Shoulder Elbow Surg. 2014;23(3):434-44.

67. Rouleau DM, Faber K, MacDermid JC. Systematic review of patient-administered shoulder functional scores on instability. J Shoulder Elbow Surg. 2010;19:1121-1128.

68. Keskula . Defining and measuring functional limitations and disability in the athletic shoulder. J Sport Rehab. 2001;10:221-231.

69. Puga VO, Lopes AD, Costa LO. Assessment of cross cultural adaptation and measurement properties of self-report outcome measures relevant to shoulder disability in Portuguese: a systematic review. Rev Bras Fisioter. 2012;16(2):85-93.

70. Hill BE, Williams G, Biolocerkowski AE. Clinimetric evaluation of questionnaires used to assess activity after traumatic brachial plexus injury in adults: a systematic review. Arch Phys Med Rehabilit. 2011;92(12):2082-2089.

71. Dorwrick AS, Gabbe BJ, Wiliiamson OD, Cameron PA. Outcome instruments for the assessment of the upper extremoity following trauma: a review. Injury. 2005;36(4):468-476.

72. The B, Reininga IH, El Moumini M, Eyngendaal D. Elbow-specific clinical rating systems: extent of established validity, reliability and responsiveness. J Shoulder Elbow Surg. 2013;22(10):1380-94.

73. Schoneveld K, Wittink H, Takken T. Clinimetric Evaluation of Measurement Tools Used in Hand Therapy to Assess Activity and Participation. J Hand Ther. 2009;22:221-235.

74. van de Ven-Stevens LA, Munneke M, Terwee CB et al. Clinimetric Properties of Instruments to Assess Activities in Patients With Hand Injury: A Systematic Review of the Literature. Arch Phys Med Rehabil. 2009;90:151-269.

75. Sambanadam SN, Priyanka P, Gul A, Ilango B. Critical analysis of outcome measures used in the assessment of carpal tunnel syndrome. Int Orthop. 2008;32:497-504.

76. Hoang-Kim A, Pegreffi F, Moroni A, Ladd A. Measuring wrist and hand function: common scales and checklists. Injury. 2011;42:253-258.

77. Bain GI, Bialocerkowski A. A systematic review of the content and quality of wrist outcome instrument. Int J Qual Health Care 2000;12:149-157.

78. Misailidou V, Mailliou P, Beneka A et al. Assessment of patients with neck pain: a review of definitions, selection criteria, and measurement tools J Chiropr Med. 2010;9:49-59

79. Schellingerhout JM, Verhagen AP, Heymans MW, et al. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. Qual Life Res 2012;21(4):659

80. de Koning CH, van den Heuvel SP, Staal JB et al. Clinimetric evaluation of active range of motion measures in patients with non-specific neck pain: a systematic review. Eur Spine J. 2008;17:905-21.

81. Holly LT, Matz PG, Anderson PA et al. Functional outcomes assessment for cervical degenerative disease. J Neurosurg Spine. 2009;11:238-244.

82. Pietrobon R., Coeytaux RR, Carey TS et al. Standard Scales for Measurement of Functional Outcome for Cervical Pain or Dysfunction. Spine (Phila Pa 1976). 2002;27:515-522.

83. Schellingerhout JM, Heymans MW, Ferhagen AP et al. Measurement properties of translated versions of neck-specific questionnaires: a systematic review.. BMC Med Res Methodol. 2011;11.

84. Gray H, Adefolarin AT, Howe TE. A systematic review of instruments for the assessment of work-related psychosocial factors (Blue Flags) in individuals with non-specific low back pain. Man Ther. 2011;16:531-543.

85. van der Meer S, Trippolini MA, van der Palen J et al. Which instruments can detect submaximal physical and functional capacity in patients with chronic nonspecific back pain? A systematic review. Spine (Phila Pa 1976). 2013;38:1608-1615.

86. Mokkink LB, Terwee CB, Startford PW et al. Evaluation of the methodological quality of health status measurement instruments. Qual Life Res. 2009;18:313-333.

87. 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.

88. Garcia-Duran Huet M, Ferrer M, Herdman M et al. BiblioPRO: online library of PRO instruments in Spanish. Qual. Life Res. 2010;19(1):62–63.

89. Nahm ML, Pieper CF, Cunningham MM .Quantifying data quality for clinical trials using electronic data capture. PLoS ONE. 2008;3(8): e3049. doi:10.1371/journal.pone.0003049

90. Porter I, Goncalves-Bradley D, Ricci-Cabello I et al. Framework and guidance for implementing patient-reported outcomes in clinical practice: Evidence, challenges and opportunities. J Comparative Effectiveness Research. 2016;5(5):507-519.

91. Black N, Burke L, Forrest CB et al. Patient-reported outcomes: pathways to better health, better services and better societies. Qual Life Res. 2016;25:1103-1112.

92. Kyte DG, Calvert M, Van der Wees PJ, Hove R, Toland S, Hille JC. An introduction to patient-reported outcome measures (PROMs) in physiotherapy. Physiotherapy. 2015;10(2):119–125.

93. Greenhalgh J, Abhyankar P, McCluskey S, Takeuchi E, Velikova G. How do doctors refer to patient-reported outcome measures (PROMS) in oncology consultations. Qual. Life Res. 2013;22(5):939–950.

94. Williams JW, Mulrow CD, Kroenke K et al. Case-finding for depression in primary care: a randomized trial. Am. J. Med. 1999;106(1):36–43.

95. 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. 20081;7(2):179–193.

96. Espallargues M, Valderas JM, Alonso J. Provision of feedback on perceived health status to health care professionals: a systematic review of its impact. Med. Care 2000;38(2);175–186.

97. Greenhalgh J, Meadows K. The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. J. Eval. Clin. Pract. 1999;5(4):401–416.

98. Marshall S, Haywood K, Fitzpatrick R. Impact of patient reported outcome measures on routine practice: a structured review. J. Eval. Clin. Pract. 2006;12(5):559–568.

99. Boyce MB, Browne JP. Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients? A systematic review. Qual. Life Res. 2013;22(9):2265–2278.

100. Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. BMJ Qual. Saf. 2014;23(6):508–518 .

101. Gonçalves Bradley DC, Gibbons C, Ricci-Cabello I et al. Routine provision of information on patient-reported outcome measures to healthcare providers and patients in clinical practice. Cochrane Database Syst. Rev. 3, 2015;CD011589.

102. Gilbody SM, House AO, Sheldon TA. Routine administered questionnaires for depression and anxiety: systematic review. Br. Med. J. 2001;322:406–409.

103. Knaup C, Koesters M, Schoefer D, Becker T, Puschner B. Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. Br. J. Psychiatry. 2009;195:15–22.

104. Khalid A. Routine administration of standardized questionnaires that assess aspects of patients' quality of life in medical oncology clinics: a systematic review. J. Egypt Natl Canc. Inst. 2013;25(2)63–70.

105. Etkind SN, Daveson BA, Kwok W et al. Capture, transfer, and feedback of patient-centered outcomes data in palliative care populations: does it make a difference? A systematic review. J. Pain Symptom Manage. 2014;49(3):611–624.

106. 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;66(3)224–241.

107. Howell D, Molloy S, Wilkinson K et al. Patient-reported outcomes in routine cancer clinical practice: a scoping review of use, impact on health outcomes, and implementation factors. Ann. Oncol. 2015;26(9):1846–1858.

108. 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. Qual Life Res. 2010;19(1):11.

109.

Figure 1: PROMs and Decision Making [19].

Figure 2: Settings for PROM collection and use [20].

Table 1: Definition and Quality Criteria for Measurement Properties [22,27,30].

Property	Description	Rating	Quality Cr
Reliability			
Internal consistency	The extent to which items in a questionnaire (sub)scale are correlated (homogeneous), thus measuring the same concept	+	(Sub)scale unidimensional ANI ≥0.70
		?	Dimensionality not known OR determined
		-	(Sub)scale not unidimensional <0.70
Reliability	The extent to which patients can be distinguished from each other, despite measurement error (relative measurement error)	+	ICC/weighted Kappa ≥0.70 OR
		?	Neither ICC/weighted Kappa, n determined
		-	ICC/weighted Kappa <0.70 OR
Measurement error	The extent to which the scores on repeated measures are close to each other (absolute measurement error)	+	MIC > SDC OR MIC outside th
		?	MIC not defined
		-	MIC ≥ SDC OR MIC equals or
Validity			
Content validity	The extent to which the domain of	+	The target population considers

Table 2: Levels of evidence for the overall quality of the measurement property [30]

Level	Rating	Criteria
Strong	+++ or ---	Consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality

Moderate	++ or --	Consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality
Limited	+ or -	One study of fair methodological quality
Conflicting	+/-	Conflicting findings
Unknown	?	Only studies of poor methodological quality

[..] Reference number, + positive result, - negative result

Table 3: ISOQOL minimum standards for patient-reported outcome (PRO) measures used in patient-centered outcomes research or comparative effectiveness research [39].

1	*Conceptual and measurement model*-A PRO measure should have documentation defining and describing the concept(s) included and the intended population(s) for use. In addition, there should be documentation of how the concept(s) are organized into a measurement model, including evidence for the dimensionality of the measure, how items relate to each measured concept, and the relationship among concepts included in the PRO measure.
2	*Reliability*-The reliability of a PRO measure should preferably be at or above 0.70 for group-level comparisons, but may be lower if appropriately justified. Reliability can be estimated using a variety of methods including internal consistency reliability, test–retest reliability, or item response theory. Each method should be justified.
3	*Validity*
3a	*Content validity*-A PRO measure should have evidence supporting its content validity, including evidence that patients and experts consider the content of the PRO measure relevant and comprehensive for the concept, population, and aim of the measurement application. This includes documentation of as follows: (1) qualitative and/or quantitative methods used to solicit and confirm attributes (i.e., concepts measured by the items) of the PRO relevant to the measurement application; (2) the characteristics of participants included in the evaluation (e.g., race/ethnicity, culture, age, gender, socio-economic status, literacy level) with an emphasis on similarities or differences with respect to the target population; and (3) justification for the recall period for the measurement application.
3b	*Construct validity*-A PRO measure should have evidence supporting its construct validity, including documentation of empirical findings that support predefined hypotheses on the expected associations among measures similar or dissimilar to the measured PRO.
3c	*Responsiveness*-A PRO measure for use in longitudinal research study should have evidence of responsiveness, including empirical evidence of changes in scores consistent with predefined hypotheses regarding changes in the measured PRO in the target population for the research application

4	*Interpretability of scores*-A PRO measure should have documentation to support interpretation of scores, including what low and high scores represent for the measured concept
5	*Translation of the PRO measure*-A PRO measure translated to one or more languages should have documentation of the methods used to translate and evaluate the PRO measure in each language. Studies should at least include evidence from qualitative methods (e.g., cognitive testing) to evaluate the translations
6	*Patient and investigator Burden*-A PRO measure must not be overly burdensome for patients or investigators. The length of the PRO measure should be considered in the context of other PRO measures included in the assessment, the frequency of PRO data collection, and the characteristics of the study population. The literacy demand of the items in the PRO measure should usually be at a 6th grade education level or lower (i.e., 12 year old or lower); however, it should be appropriately justified for the context of the proposed application

Table 4: Criteria to consider when evaluating a systematic review of measurement properties of a PROM[1].

Criteria	**Description**
Question	Was the question clearly articulated and understandable? [i.e., was the specific subject matter area to which the PROMs apply (e.g., rotator cuff disease) clear from the question]
Search Strategy	Was the search strategy completely described and appropriate? Is it unlikely that the search methods could have resulted in missed primary studies? (i.e., were appropriate databases and search terms used, was there an attempt to hand search or review reference lists and contact experts in the area to identify additional studies)
Article Selection	Were there explicit and appropriate criteria for inclusion and exclusion of the primary studies (did these follow directly from the question posed)? Were there two individuals who separately and independently performed the article selection? Was agreement reported and was it sufficient? Were the number of articles excluded at each stage and were the reasons for exclusion reported (i.e. was the exclusion biased in any way)?
Data Extraction	Were there two individuals who separately and independently performed the data extraction? Was there training to do so appropriate? Was agreement reported and was it sufficient?
Methodological Quality/Risk of Bias Assessment	Were explicit criteria reported and used to assess the methodological quality of the included studies? Were the criteria appropriate for the measurement property(ies) assessed (e.g., COSMIN criteria)[23,27-28]?
Measurement property assessment	Were explicit criteria reported and used to assess the measurement properties of the included studies? Were the criteria appropriate for the measurement property(ies) assessed [22]?
Synthesis Methods	Were explicit criteria reported and used to synthesize the methodological quality and the measurement properties of the included studies? Were the criteria appropriate [e.g., 30-32]?
Heterogeneity	Did the authors account for differences between the primary studies evaluating the same instruments (e.g., study setting, sample characteristics, study methods)?
Publication Bias	Did the authors attempt to assess publication bias using funnel plots where applicable (e.g., for quantitative criteria)?

Other	Were other possible sources of bias in the systematic review avoided? (e.g., funding bias[2], investigator bias[3])

1. If the answer is "no" to any of these questions, the methods of the review are flawed which could bias the results. The degree that risks to bias actually influence the results is determine by a rationale assessment of overall study methods and results.
2. Funding bias may result from a funding body supporting the systematic review whom also own the copyright or intellectual property of some PROM under review.
3. Investigator bias may arise from authors of the systematic review also being authors of included primary studies or creators of the included PROM(s).

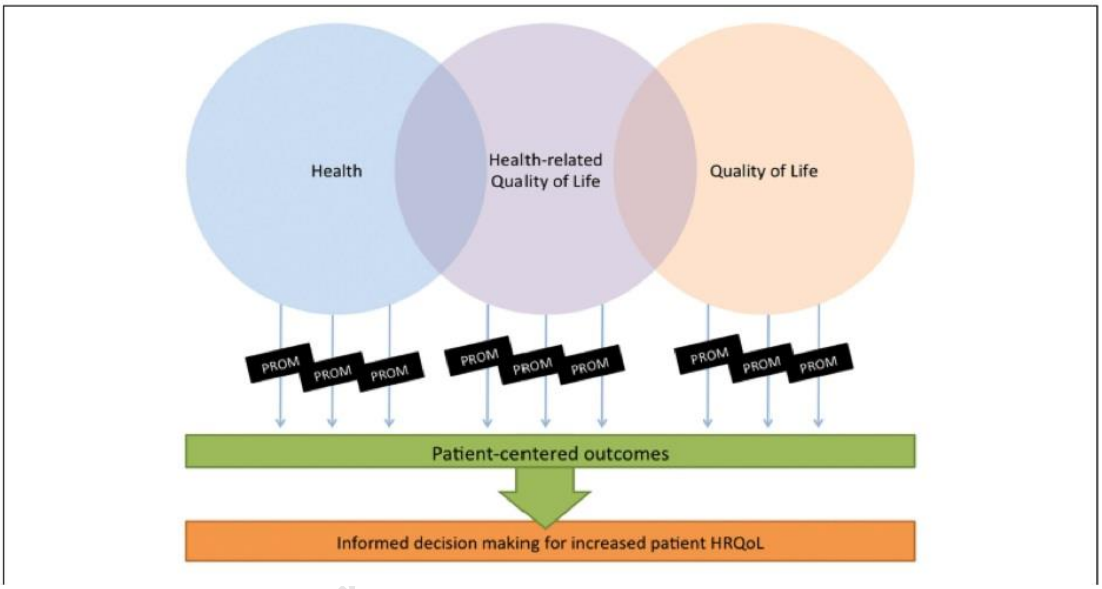

Figure 1

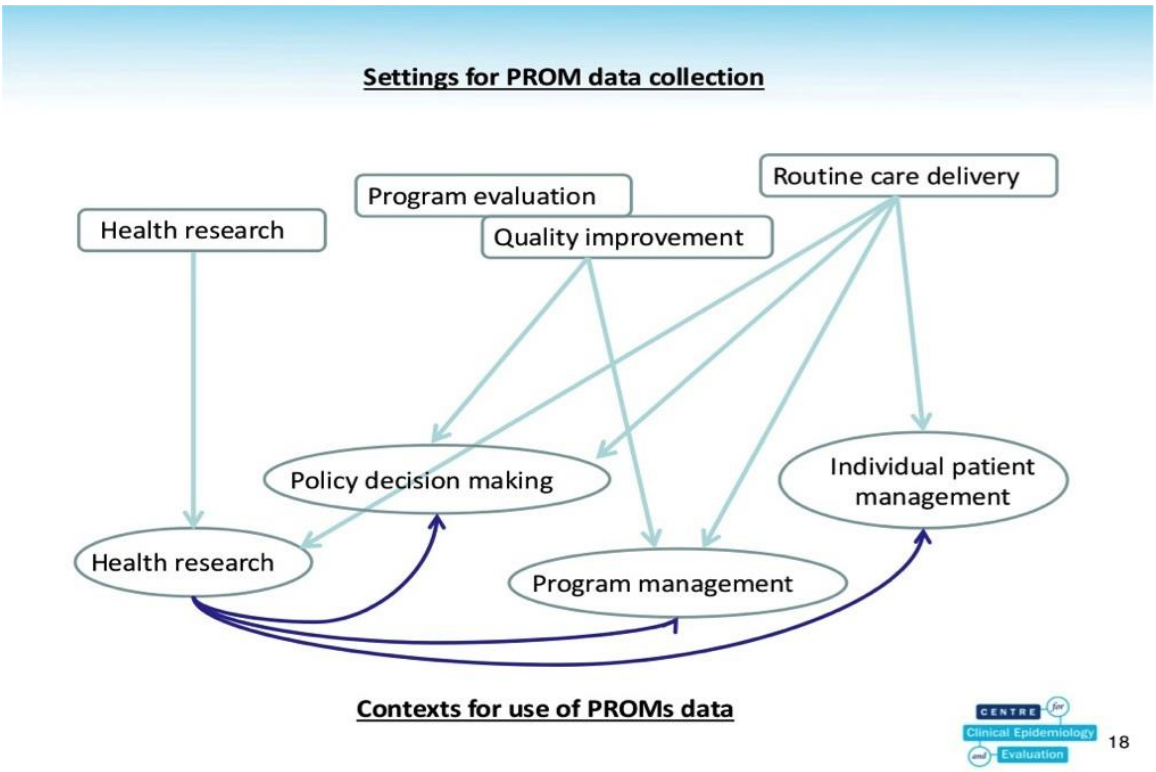

Figure 2