

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



Browne, JP; Cano, SJ; Smith, S (2017) Using Patient-reported Outcome Measures to Improve Health Care: Time for a New Approach. Medical care. ISSN 0025-7079 DOI: <https://doi.org/10.1097/MLR.0000000000000792>

Downloaded from: <http://researchonline.lshtm.ac.uk/4329137/>

DOI: [10.1097/MLR.0000000000000792](https://doi.org/10.1097/MLR.0000000000000792)

Usage Guidelines

Please refer to usage guidelines at <http://researchonline.lshtm.ac.uk/policies.html> or alternatively contact researchonline@lshtm.ac.uk.

Available under license: <http://creativecommons.org/licenses/by-nc-nd/2.5/>

Using patient-reported outcome measures to improve healthcare: time for a new approach.

Key words: Patient-reported outcome measures; quality of life; quality improvement; feedback; value-based purchasing.

Patient-reported outcome measures (PROMs) are measures of health, quality of life and related constructs that come directly from patients. The importance of PROMs in demonstrating the benefit to patients of new health technologies is widely recognised and they have been endorsed by the UK's National Institute for Health Care Excellence and the Food and Drug Administration in the United States. PROMs succeed within well conducted studies of effectiveness because it is clear how to use the information they provide.

There is increasing interest in embedding PROMs within health information systems to compare healthcare providers and it has been claimed that this has the potential to transform healthcare into a more patient-centred model.¹⁻³ PROMs have inherent advantages over alternative performance measures such as mortality and clinician-defined morbidity. They tap into the unique and authoritative insight of patients about their own health, avoid the conflict of interest that occurs when healthcare providers rate their own performance, and directly measure the extent to which the main objective of most interventions has been achieved: improving the patient's health and quality of life. The routine collection of PROMs has already been implemented in many countries², and it is clear they should have a role in healthcare intelligence. But what is that role? We contend that previous attempts to use routinely collected PROMs have disappointed and propose a shift in purpose. We focus on the use of routinely collected PROMs to compare healthcare providers and do not question the value of PROMs within comparative effectiveness research. We also do question the use of patient-reported experience measures, sometimes referred to as 'PREMS', to compare providers.

When PROMs are embedded within health information systems they are predominantly used to compare the performance of individual service providers,⁴ networks of service providers,⁵ or insurance providers.⁶ There are two mechanisms by which these comparisons are thought to improve healthcare. First, 'value-based purchasing', where PROMs are used to facilitate purchaser choices and encourage competition.⁷ Second, 'audit and feedback' where PROMs are used to uncover flaws in care processes and competencies.⁸

In the United States PROMs have been used as value-based purchasing tools within the 'star ratings' of Medicare Advantage insurance plans. These influence reimbursement, physician bonuses and the ability of insurance plans to expand.⁶ In England PROMs have been used to compare NHS Trust performance for four common surgical procedures since 2011 and are also linked to financial rewards.⁴ The evidence base supporting these initiatives is weak and there has been strong resistance from experts in performance assessment to comparing organisations with PROMs.⁹ For example, an evaluation of the English PROMs Program has found no impact on patient outcomes¹⁰ and the whole Program is now under review.⁴

PROMs were not originally developed to compare providers and they lack many of the attributes desirable in a performance indicator. Variation in PROMs is heavily influenced by patient level variables because health perceptions are influenced by factors other than the healthcare provider where treatment was received. Typically, less than 10% of the variation in PROMs is at the provider level,¹⁰ a threshold often considered the minimum requirement for a performance indicator.¹¹ High patient-level heterogeneity reduces the precision of provider estimates, which in turn increases the sample sizes and time needed to detect clinically meaningful variation.¹¹ There is evidence in some fields such as bariatric surgery that slightly more than 10% of variation in PROMs can be explained at the provider level but

even here the amount of variation that can be explained by patient-level factors is much higher.¹² An important source of patient-level variation is the type of treatment received. In breast reconstruction after mastectomy, for example, patients undergoing autologous tissue procedures report better outcomes compared to patients who received implants alone.¹³ This patient-level variation occurs within providers because breast cancer surgeons offer a range of options to their patients. It can be argued that surgeons should not offer implant-only procedures to their patients but this ignores the role that the patients' personal values for different outcomes play in the decision-making process. Over time, as evidence on the comparative effectiveness of autologous procedures emerges it is possible that a greater proportion of women will opt for this type of reconstruction, but this will happen because of analyses that quantify the value of different types of surgery, as opposed to different providers. Other important patient-level influences on PROMs that are beyond the control of the original healthcare provider include new healthcare episodes, life events and health behaviors such as poor adherence to rehabilitation plans. As the time from index episode to outcome assessment passes the influence of the original healthcare provider on outcome weakens and the influence of these factors grows. These factors are unlikely to be randomly distributed across providers and cannot be controlled for by case-mix adjustment.

A broader concern relates to the effectiveness of value-based purchasing as a whole. A 2013 review of systematic reviews found that the evidence supporting its use was not convincing, particularly when outcome, rather than process measures were used to quantify value.¹⁴ A key difficulty with value-based purchasing is the low propensity of purchasers, including patients¹⁵, and agents acting on their behalf¹⁶, to use publicly released healthcare performance data when selecting providers. In 2016, a taskforce which reviewed the NHS

PROMs Program concluded that patients and clinicians may not appreciate the initiative because data is not easily accessible or presented in ways that demonstrate relevance to routine care choices.⁴

PROMs have also been used to stimulate quality improvement through the 'audit and feedback' model. Here, PROMs are aggregated to provider level and compared, but the data are not necessarily released to the public. Poor performers are hypothesized to have an intrinsic motivation to improve relative to their peers.¹⁷ This can be done through local investigation of flaws in care processes and competencies or by learning from providers with the best results.^{18,19} To date there is little evidence to support this use of PROMs.^{20,21} A 2013 systematic review found only one study had evaluated the benchmarking of healthcare providers using PROMs and this found no evidence of effectiveness.²² The difficulty is not the audit and feedback model itself, which, unlike value-based purchasing, has a solid evidence base.²³ The difficulty lies in the use of PROMs in isolation, which violates the core principles of audit and feedback. The model involves providing a recipient with a summary of performance over a defined time period, focusing on evidence-based practices.²⁴ Crucially, the feedback should relate to behavior, for example, hand hygiene practices, as opposed to the outcome of those practices. PROMs are outcome measures: they tell us whether a patient has had a good or bad outcome, not what caused that outcome. There is nothing in a PROM that identifies the flaws in care processes that led to poor outcome – the measures focus on patients' perceptions about their health and quality of life rather than on the care they received. There is now a widespread agreement that audit and feedback only works when it identifies specific individuals or groups responsible for poor performance, targets specific behaviors, is based on recent performance and produces a clear action plan

for improvement.²⁴ PROMs, when used in isolation, do not lend themselves to these specifications. They can only be used to improve the quality of individual healthcare providers when linked to data about their care processes and structures. To date this has proved very difficult and there are few credible success stories.²⁵ This is due to the logistics of data linkage and the complexity of the causal pathway that determines patient level outcomes.²⁵ In hip replacement, for example, there are many different organisation-level variables that might influence pain and function for an individual patient six months after surgery. Collecting only PROMs data would not help us understand whether differences in PROMs were due to pre-operative assessment arrangements, surgeon ability, rehabilitation protocols or access to post-discharge facilities. Also, PROMs take a long time to collect and analyse because of the need to wait until the full effects of a treatment have been realized.²⁵ In hip and knee replacement, for example, PROMs cannot be collected until a minimum of six months after surgery because the full benefits of surgery are difficult to judge any earlier.²⁶ Feedback to orthopedic surgeons is usually at least one year out of date by the time it is reported. This is not a reasonable length of time for an outcome measure to be used to improve performance given the amount of natural change in staffing and practice that may happen in the interim. The flaws described above are recognized by clinicians and are a significant impediment to the local use of PROMs in quality improvement activities.²⁷ The flaws are not insurmountable and it is possible that initiatives with rapid feedback, comprehensive data collection and analyses that tease apart the relationships between local processes and outcomes, could lead to concrete and effective quality improvement solutions for healthcare providers. This will require considerable investment however, and a willingness to move beyond relying solely on the motivations of clinicians and providers.

Clearly there are substantial flaws in current approaches to the use of routinely collected PROMs. There are other ways we could use PROMs at both the macro and micro level which have a greater probability of leading to sustained improvements in health care (Table 1). Previously these have been considered ‘secondary uses’ of routinely collected PROMs. Our contention is that these uses have a much greater likelihood of improving healthcare for whole patient groups than using PROMs as performance measures.

First, it is possible to use PROMs to quantify the independent burden of different conditions once comorbidity has been accounted for.²⁸ This in turn allows us to identify conditions where technological advances would lead to the greatest benefit for patients, and where research funding should be prioritized. By linking PROMs to data on care access and insurance coverage it is also possible to identify the alleviation in burden that could be achieved by improving access to current technologies.²⁸

Second, PROMs can help us assess the real-world impact of new technologies as they diffuse through care systems. Such technologies are often introduced based on weak evidence, or based on research that is not replicated in practice.²⁹ Routinely collected PROMs allow for a broad research program across different conditions and interventions and help to identify where technological advances are an advance over standard treatments. The Swedish Hip Arthroplasty Register and the Veteran’s Health Study in the United States are models for how PROMs can be used to perform comparative effectiveness research within large observational datasets.^{18,30}

Third, routinely collected PROMs could be used to assess the effectiveness of major policy initiatives where RCT level evidence is not feasible. PROMs data have been used to demonstrate the impact of introducing independent providers in the English NHS³¹, the effects of introducing market competition between hospitals,³² and to compare the Veteran's Health System with private managed care in the US.³³

Fourth, PROMs data can be used to identify conditions or interventions where substantial variation in patient outcome exists across providers. Rather than assume that variation exists it makes more sense to check first, and then use this information to guide where further, more intensive quality improvement initiatives should take place. The English NHS PROMs Program focuses on four high volume surgical procedures which all have low outcome variation at the provider level.¹⁰ Volume is an important consideration when developing quality improvement projects but should be secondary to evidence, as with bariatric surgery, that substantial provider level variation exists.

Finally, PROMs can improve the choice of treatment plans for individual patients. The evidence-base supporting the use of PROMs at the individual patient level is weak.²² But the PROMs used in this literature were developed for group level analysis using an approach known as 'Classical Test Theory', a methodology not designed to develop measures for use at the individual patient-level. These PROMs are not reliable enough to provide a precise estimate of a patient's current health status,³⁴ and they do not have the interval-level properties necessary to interpret the meaning of change at different points on the scale.³⁵ Instead, newer methods designed to track the progress of individual patients, using approaches such as Rasch Measurement Theory,³⁶ or patient-reported symptom checklists³⁷

should be adopted on a more widespread basis. These newer approaches allow clinicians to track treatment progress and tailor therapies appropriately. When used in conjunction with other clinical information they can be used to highlight instances where individual patients are unlikely to derive benefit from a treatment or where current treatment is ineffective. There is emerging evidence in cancer care that this approach can improve both the process and outcome of care.³⁷

The suggestions above will require a significant investment in data collection technology and training. At a macro level, they imply the routine and regular collection of PROMs with all patients who interact with a health system, not just small sub-groups on a once-off basis. Professionals and administrative staff will need practical and methodological support, including training, to ensure that data is collected and used properly. At present, many initiatives have assumed that healthcare providers will find the staff and materials to collect PROMs data from their own resources. There will also need to be significant investment in the development and standardization of electronic health record systems to accommodate PROMs.

At a micro level, we recommend an organic, bottom-up approach to the adoption of PROMs. This is already happening in specific disciplines such as breast cancer surgery³⁸ but there are many areas where PROMs are rarely used at the individual patient level, or the wrong type of PROMs are used. Policy makers should encourage the gradual adoption of modern PROMs through pilot projects in collaboration with psychometric experts rather than impose a standardized approach.

Our core message is that the existing focus on using routinely collected PROMs to compare providers is unlikely to produce benefits for patients and needs to change. We are not suggesting that such comparisons will never be useful: it is possible that PROMs may eventually become a useful way to benchmark healthcare providers in areas with substantial provider-level variation, and where other information about the processes of care can be easily linked to PROMs to pinpoint the causes of poor performance. Until these conditions are met the use of PROMs to compare providers should be treated as a new technology with little evidence of effectiveness and the potential to be harmful. National performance measurement systems can lead to many unintended and harmful consequences such as inequality of access, erosion of trust and damaged staff morale.³⁹ Such systems should be introduced very carefully and only when the benefits outweigh the harms. We urge policy makers to continue with initiatives such as the English NHS Programme but to shift the focus of analysis away from provider comparisons. At present, we believe it would be more useful to focus on other sources of variation if the full potential of routinely collected PROMs is to be realized.

References

- 1 Devlin N, Appleby J. Getting the most out of PROMs: putting health outcomes at the heart of NHS decision-making. London (UK): King's Fund; 2010.
- 2 Recommendations to OECD Ministers of Health from the high level reflection group on the future of health statistics. Strengthening the international comparison of health system performance through patient-reported indicators. Paris: Organisation for Economic Co-operation and Development; 2017.
- 3 Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013;346:f167.
- 4 Kyte D, Cockwell P, Lencioni M, et al. Reflections on the national patient-reported outcomes programme (PROMs). Where do we go from here? *J R Soc Med* 2016;109:441-445.
- 5 Wright SM, Craig T, Campbell S, et al. Patient satisfaction of female and male users of Veterans Health Administration Services. *J Gen Intern Med* 2006;21(suppl 3):S26–32.
- 6 Cotton P, Newhouse JP, Volpp KG, et al. Medicare Advantage: issues, insights and implications for the future. *Popul Health Manag* 2016;19(Suppl 3):S1–S8.
- 7 Porter M. What is value in health care? *N Engl J Med* 2010;363:2477-2481.
- 8 Bickman L. A measurement feedback system is necessary to improve mental health outcomes. *J Am Acad Child Adolesc Psychiatry* 2008;47:1114–1119.
- 9 Lilford R, Mohammed M, Spiegelhalter D, et al. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004;363:1147-1154.

- 10 Varagunam M, Hutchings A, Neuburger J, et al. Impact on hospital performance of introducing routine patient reported outcome measures in surgery. *J Health Serv Res Policy* 2014;19:77-84.
- 11 Fung V, Schmittiel JA, Fireman B et al. Meaningful variation in performance: a systematic literature review. *Med Care* 2010;48:140–148.
- 12 Waljee JF, Ghaferi A, Finks JF et al. Variation in patient-reported outcomes across hospitals following surgery. *Med Care* 2015;53:960-6.
- 13 Jeevan R, Browne JP, Gulliver-Clarke C et al. Surgical determinants of patient-reported outcomes following post-mastectomy reconstruction in women with breast cancer. *Plast Recon Surg* 2017;139:1036-1045.
- 14 Eijkenaar F, Emmert M, Scheppach M, et al. Effects of pay for performance in health care: a systematic review of systematic reviews. *Health Policy* 2013;110:115– 130.
- 15 Shekelle P, Lim Y-W, Mattke S, et al. Does public release of performance results improve quality of care. A systematic review. London: The Health Foundation; 2008.
- 16 Allen P. An economic analysis of the limits of market based reforms in the English NHS. *BMC Health Serv Res* 2013;13(Suppl 1):S1.
- 17 Moriarty JP, Smallman C. En route to a theory of benchmarking. *Benchmarking Int J* 2009;16:484–503.
- 18 Lindgren JV, Wretenberg P, Karrholm J, et al. Patient-reported outcome is influenced by surgical approach in total hip replacement: a study of the Swedish Hip Arthroplasty Register including 42,233 patients. *Bone Joint J* 2014;96-B:590-6.

- 19 Stromqvist B, Fritzell P, Hagg O, et al. The Swedish Spine Register: development, design and utility. *Eur Spine J* 2009;18(Suppl 3):294–304.
- 20 Weingarten SR, Kim CS, Stone EG, Kristopaitis, et al. Can peer-comparison feedback improve patient functional status? *Am J Managed Care* 2000; 6:35–39.
- 21 Boyce M, Browne JP. The effectiveness of providing peer benchmarked feedback to hip replacement surgeons based on patient-reported outcome measures—results from the PROFILE (Patient-Reported Outcomes: Feedback Interpretation and Learning Experiment) trial: a cluster randomised controlled study. *BMJ Open* 2015;5:e008325.
- 22 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;22:2265-78.
- 23 Ivers NM, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;6:CD000259.
- 24 Ivers NM, Sales A, Colquhoun H, et al. No more ‘business as usual’ with audit and feedback interventions: towards an agenda for a reinvigorated intervention. *Implement Sci* 2014;9:14.
- 25 Greenhalgh J, Dalkin S, Gooding K, et al. Functionality and feedback: a realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care. *Health Serv Deliv Res* 2017;5:2.
- 26 Browne JP, Bastaki H, Dawson J. What is the optimal time point to assess patient-reported recovery after hip and knee replacement? A systematic review and analysis of routinely reported outcome data from the English Patient-Reported Outcome Measures Programme. *Health Qual Life Out* 2013;11:128.

- 27 Boyce M, Browne JP, Greenhalgh J. The experiences of professionals who use patient-reported outcome measures as quality improvement tools - a systematic review of qualitative research. *BMJ Qual Saf* 2014;23:508-18.
- 28 Sullivan PW, Ghushchyan V. Preference-Based EQ-5D Index Scores for Chronic Conditions in the United States. *Med Decis Making* 2006;26:410–420.
- 29 Rothwell PM. External validity of randomised controlled trials: to whom do the results of this trial apply? *Lancet* 2005;365:82-93.
- 30 Hamad A, Lee A, Ren XS, et al. Use of antidepressant medications: are there differences in psychiatric visits among patient treatments in the Veterans Administration? *Med Care* 2004;42:551-559.
- 31 Chard J, Kuczawski M, Black N, van der Meulen J; POiS Audit Steering Committee. Outcomes of elective surgery undertaken in independent sector treatment centres and NHS providers in England: audit of patient outcomes in surgery. *BMJ* 2011;343:d6404.
- 32 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 Serv Res Policy* 2015;20:11-7.
- 33 Selim AJ, Berlowitz D, Kazis LE, et al. Comparison of health outcomes for male seniors in the Veterans Health Administration and Medicare Advantage plans. *Health Serv Res* 2010;46:376-96.
- 34 Browne JP, Jamieson L, Lewsey J et al. Patient-Reported Outcome Measures (PROMs) in Routine Elective Surgery. Report to the Department of Health. London: London School of Hygiene and Tropical Medicine. 2007.

- 35 Palta M, Chen HY, Kaplan RM, et al. Standard error of measurement of five health utility indexes across the range of health for use in estimating reliability and responsiveness. *Med Decis Making* 2011;31:260–269.
- 36 Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurol* 2007;6:1094–105.
- 37 Basch E, Deal AM, Kris MG, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. *J Clin Oncol* 2016;34:557-65.
- 38 Cohen W, Mundy L, Ballard T, et al. The BREAST-Q in surgical research: a review of the literature 2009-2015. *J Plast Recon Aesth Surg* 2016;69:149-62.
- 39 Mannion R, Braithwaite J. Unintended consequences of performance assessment in healthcare: 20 salutary lessons from the English National Health Service. *Intern Med J* 2012;42;569-574.