Why it is hard to use PROMs and PREMs in routine health and care

Tim Benson   [1,2]

¹R-Outcomes Ltd, Newbury, UK
²Institute of Health Informatics, UCL, London, UK

Correspondence to
Tim Benson;
tim.benson@r-outcomes.com

ABSTRACT

Patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs) show the results of healthcare activities as rated by patients and others. Patients or their proxies record feedback using questionnaires. These can enhance quality for all and tailored care for individuals. This paper describes obstacles that inhibit widespread use of PROMs and PREMs and some potential solutions.

Implementation is a prerequisite for any innovation to succeed. Health and care services are complex and people need to be engaged at every level. Most people are cautious about proven innovations such as PROMs and PREMs but champions and leaders can help them engage. The NASSS framework (reasons for Non-adoption, Abandonment and failure to Scale up, Spread or Sustain digital health innovations) helps indicate that implementation is complex why it may be resisted. The Plan-Do-Study-Act (PDSA) approach aids implementation and helps ensure that everyone knows who should do what, when, where, how and why. Noise is an under-appreciated problem, especially when tracking patients over time such as before and after treatment. Interoperability of PROMs and PREMs with electronic health records should use Fast Health Interoperability Resources and internationally accepted coding schemes such as SNOMED CT and LOINC.

Most projects need multiple measures to meet the needs of everyone involved. Measure selection should focus on their relevance, ease of use, and response rates.

If these problems are avoided or mitigated, PROMs and PREMs can help deliver better patient outcomes, patient experience, staff satisfaction and health equity.

PATIENT-REPORTED OUTCOME AND EXPERIENCE MEASURES

The objectives of health and care activities are both their direct benefit to patients, such as improved health outcomes and experience, and the effective use of financial resources.[1] Our focus here is on why measures of patient outcomes and experience are not used routinely in all health services and how we can address the barriers to use.

We cannot improve what we do not measure. High-quality healthcare is care that is effective, safe and provides as positive experience as possible. The impact on patients can be measured by patient-reported outcome measures (PROMs); specific services can be assessed by patient-reported experience measures (PREMs).[2]

PROMs indicate the outcome of care from the patient's point of view. They assess whether health status and well-being have changed and if treatment goals have been met and also other patient-specific aspects such as social determinants of health. Patients are usually the subjects of PROMs but other people, such as carers and staff, may be subjects sometimes. The reporter is usually the patient, but a proxy such as a family member, carer or member of staff may stand in if the patient cannot respond due to factors such as age, health conditions or language.[3] An identifier is needed to track change over time, or before and after treatment; this should be linked to the patient's electronic health record (EHR), to help clinicians to tailor care to each patient's needs.[4]

PREMs measure a healthcare provider's services from the patient's viewpoint. The service provider (eg, hospital or clinic) is usually the subject. The reporter is the patient in most cases, although a relative, carer, or staff member may act as proxy. PREMs are usually anonymous due to the potential sensitivity of feedback.[5]

PROMs and PREMs fall into two broad classes, generic or specific. Generic measures cover all patients regardless of diagnosis or treatment, while specific measures cover patients a single diagnosis or treatment. Specific PROMs were often developed and used by pharmaceutical companies for use in clinical trials to obtain regulatory approval.[6] Length and ease of use are less critical in clinical trials than in regular healthcare settings.

PROMs and PREMs generate numeric (quantitative) feedback from patients, staff and carers but may also include free-text (qualitative) questions, where respondents express their thoughts in their own words.

National surveys show that satisfaction with health and social care services is at an all-time low. In 2022, overall satisfaction was 29% with the NHS and 14% with social care services.[7] These results demand action.

Survey response rates are also disappointing. The English General Practice Patients' Survey is sent to over 2 million people each year. The 2022 results show a response rate under 30%, after five reminders, 14 language versions and allowing either online or paper response[8] The NHS Friends and Family Test has one standard question plus free text. For July 2023, the national response rates were 22% for in-patients and 11% for accident and emergency department attendees.[9]

IMPLEMENTATION

Innovations, however good, are of no value unless implemented and used. PROMs and PREMs face similar barriers to implementation as other health innovations. These include the complexity of the healthcare system where many groups and people have the power to prevent or inhibit adoption; the problem of noise, which is often unrecognised; the difficulty of sharing data (interoperability), and inadequate planning and leadership.

This paper adresses some of the main barriers which impact the implementation of PROMs and PREMs. Implementation science focuses on how best to introduce innovations.[10] It aims to close the gap between what is known to work and what people do (the *know-do gap*). Implementation is a pre-condition for any innovation to spread.[11] Implementation is different from innovation. In healthcare, there is often a long time gap between successful demonstration of an innovation and its widespread use. For example, EHRs were first implemented in hospitals 50 years ago, but they have only become common since 2010.[12]

The two most widely used generic PROMs are the Short Form SF-36[13] and the EuroQol EQ-5D,[14] which were developed during the 1980s. The most widely used PREM is the Hospital Consumer Assessment of Health Care Providers and Systems,[15] which was developed during the 2000s.

The Consolidated Framework for Implementation Research (CFIR) was designed to help implementers. The original CFIR (2009) had 5 major domains and 39 constructs.[16] The latest version (2022) has grown to 67 constructs:[17]

► Innovation (8 constructs)—what is being implemented.
► Outer setting (10 constructs)—the broader context (may contain multiple outer settings).
► Inner setting (21 constructs)—where the innovation is being implemented.
► Individuals (13 constructs)—the individuals in the project.
► Implementation process (15 constructs).

For outcomes evaluation, several frameworks have been proposed including the CFIR Outcomes Addendum,[18] Reach, Effectiveness, Adoption, Implementation Maintenance,[19] the Implementation Outcomes Framework[20] and reasons for Non-adoption, Abandonment and failure to Scale-up, Spread and Sustain (NASSS).[21] Our own work has used NASSS (see below).

It is always hard to introduce innovations into any healthcare environment. A few people are keen to embrace innovation, but most are cautious. Rogers groups people in any population into five categories:[22]

► Innovator: 2.5%
► Early adopter: 13.5%
► Early majority: 34%
► Late majority: 34%
► Laggard: 16%.

Most decisions in healthcare are made in committees where innovators and early adopters are a minority. Advocates of PROMs and PREMs must show how they provide benefits to people at all levels (patients, clinicians, managers and payers). Champions and leaders often play a critical role in winning acceptance.[23]

In the next sections, we consider some issues that inhibit PROMs and PREMs implementation, including complexity (using the NASSS Framework), noise, innovativeness and interoperability. Finally, we address some specific issues of using PROMs and PREMs, such as questionnaire design and the range of measures needed.

Complexity

Complexity theory distinguishes between systems that are simple, complicated or complex.

► *Simple* systems show straight-forward cause and effect behaviour. Hitting a ball with a bat is a simple system.
► *Complicated* systems have many more parts but each part follows cause and effect. Sending a rocket to the moon is a complicated system.
► *Complex* systems are hard to model and the outcomes can be unpredictable. Healthcare systems are usually complex.[24]

The NASSS Framework was originally conceived to understand why many healthcare IT systems failed to achieve their objectives.[21] NASSS has seven levels and uses the lens of complexity theory. This section shows the original NASSS name for each level, plus short comments about its relevance to PROMs and PREMs.

1. *The condition or illness.* The number of possible clinical conditions is very large, although in some specialist units, such as oncology and maternity, most patients have a single known condition. In others patients may have multiple diagnoses, or the diagnosis may not be known. When this is the case, generic measures are more appropriate.
2. *The technology.* The technology includes the questionnaire, data collection and reporting tools. Survey designers need to be clear about what they aim to achieve, what to measure and the end-to-end process of planning, data collection, analysis and use of results. The measures selected need to have been validated.
3. *The value proposition* for the supplier, purchaser and users differs. Some purchasers are fixated about choosing free measures, without understanding that these come with little or no support, training or software. Questionnaires, measures or the whole service may be charged at a flat rate (all you can eat), per response

or be free. Purchasers must always consider the cost of the whole process, including hardware and software, license fees, support, training, and results analysis.

4. *The adopter system* comprises professional staff, patients and lay caregivers. At each stage in the process, it is important to be clear about who does what, when, where, how and why. Our experience is that patients will complete questionnaires if asked by people they respect (eg, their clinicians) and if the terms used are familiar. Also, staff will do what they are asked if the request comes from a line manager to whom they report. However, if patients or staff are not engaged, poor response rates are to be expected.

5. *The organisation(s)*. Managers and leaders can motivate staff, provide resources and remove or mitigate barriers to use. They can ensure that the results are used to improve efficiency and the quality of care for all.

6. *The wider context* (institutional and societal). Most healthcare systems are managed top-down and people usually try to do what they are asked to. Sometimes, people at national or regional bodies try to pick winners and mandate their use, but this can have mixed success.

7. *Interaction and mutual adaptation over time.* As people, policies and technology change, so must the process. For example, a decade ago most questionnaires were completed on paper, but today most are online.

When all NASSS domains are simple (which is rare), the programme is likely to be easy to implement, on time and within budget. When many domains are complicated, the programme is achievable, but it will be difficult and likely to exceed both timescale and budget. When domains are complex, the chances of success are low. In all cases, the aim should be to work with the system, reduce complexity and avoid problems.[25]

Plan-Do-Study-Act

The Plan-Do-Study-Act (PDSA) approach helps implementers of any innovation think about who must be engaged at every stage.[26] PDSA includes three questions to answer before starting:
► What are we trying to achieve?
► How will we measure improvement?
► What changes will we do?
The stages in the PDSA cycle are:

1. *Plan.* Staff at all levels and stages must be involved early on, so people know who is to do what, when, where, how and why. Support from a recognised leader or champion is invaluable, to make sure that everyone has the tools, training, time and resources required.

2. *Do.* Ultimately, responses are analysed by computer. Responses may be collected digitally or on paper, although paper responses have to be entered into a computer later. Ideally collect digitally at source.

3. *Study.* Examine results quickly and disseminate widely. Use thresholds as simple comparators. Using 0–100 scales, we use: under 40 bad; 40–59 poor, 60–79 moderate; over 80 good.

4. *Act.* Trust the results and act as quickly as possible. What you do depends on local circumstances, but do not ignore the findings. Our experience suggests that unexpected results are usually correct.

PDSA cycles usually run sequentially. This is helpful if a cycle suggests that the approach needs to be revised. PDSA cycles may also be run simultaneously, but this is harder to manage when changes are required.

Response rate is a success indicator which is easy to measure.[27] A poor response rate usually means that potential respondents were not asked, or that they deemed the questionnaire to be irrelevant or burdensome. The following response rate bands are helpful: under 10% bad; 10–30% poor; 30–60% moderate; over 60% good.

Noise

Noise is scatter; it is always present in surveys.[28] Noise is a serious problem, especially when the main result is the difference between two responses (two noisy numbers) for the same person (eg, before and after treatment).

There are three main types of noise[29]:
► *Level noise* happens when respondents are consistent in their differences. For example, some people always avoid the endpoints of scales, but others do not. Level noise also occurs when some respondents differ in their understanding of a question.
► *Stable pattern noise* happens when respondents prioritise different aspects of the problem. For example, some people prioritise physical symptoms and some mental symptoms.
► *Occasion noise* occurs when scores vary in ways that are hard to predict in advance. For example, some people respond differently when rushed or hungry.

One way to reduce noise is to limit the number of options. For example, most people are consistent about whether they agree or not with a statement. They can also say whether they agree strongly or weakly. Being non-committal does not mean agreement—it may just be a polite way of disagreeing. A short scale (eg, 4 or 5 options) creates less noise than a long scale (ten or more options).[30] Always think about how you expect people to rate each question and aim for a broad spread of answers representing different views. It says little if everyone answers the same way.

When designing a survey, decompose complex, multidimensional judgements by taking one dimension at a time. It is hard to compare across different dimensions. Also, put overall assessments near the end, not at the start of a questionnaire.[31]

Interoperability

PROMs need to be interoperable with EHRs to help clinicians tailor care for individual patients. For example, in a randomised controlled trial of patients having chemotherapy, adding the use of PROMs between clinic visits (importing results into the EHR, reviewing and acting on the findings) led to an improvement in 5-year survival from 33% to 41%.[32]

There are two primary use cases for interoperability: (1) the clinician asks the survey engine to administer a specific questionnaire; (2) the EHR imports results data from the survey engine for use by clinicians and ideally patients.

Fast Health Interoperability Resources (FHIR) Release 4 (or later) is now the preferred way to do interoperability in most countries (including USA and UK). FHIR is built around resources. FHIR has two special resources for use with surveys: *Questionnaire*, which specifies the questionnaire itself, and *QuestionnaireResponse*, which specifies the answers of a specific subject (patient). *QuestionnaireResponse* is needed for most clinical purposes, along with resources to identify the patent, clinic and so on.

All fields in a resource are optional. Resources must be simplified (profiled) and combined with other resource profiles to create an Implementation Guide (IG) for each transaction. The best IGs have the minimum optionality and define the exact format, including what codes shall be used. Implementers use IGs, not the base standard.[33]

Information exchanged must be unambiguous, which means that concepts must be coded. The two main coding schemes used in many countries are the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) and Logical Observation Identifiers Names and Codes (LOINC). In 2022, it was agreed that all LOINC codes would have an exact SNOMED CT equivalent.[34] Some of the concepts used in questionnaires have the codes required, but not all.

INNOVATION
Questionnaire design
People will complete questionnaires when the following conditions are met:
1. They are asked by someone they respect.
2. They believe it will help themselves or others.
3. The questionnaires are relevant, quick and easy to complete.

When designing a questionnaire, think about the benefits to those who need to act on the results as well as to the respondents who complete them.

Scoring creates problems for people using the results when different measures use different scales, endpoints or directions. For example, the NHS PROMs scheme for Hip Replacement surgery uses measures with three

Table 1 Short descriptions of the range of measures that may be used

Domain	Type	Measures (items)
How You Feel	PROM	1. Health status (physical and mental symptoms, disability, dependence) 2. Personal well-being (life satisfaction, worthwhile, happiness, anxiety) 3. Person-specific measures (two key issues) 4. Sleep (sleep and wake at same time, refreshed, sleep quality) 5. Fatigue (energy, tire fast, concentrate, stamina).
Impact of Care	PROM	1. Health confidence (health literacy, self-efficacy, access to help, shared decision-making) 2. Self-care (diet, physical activity, weight, medication) 3. Shared decisions (benefits, downside, choices, involved) 4. Behaviour change (capability, opportunity, motivation) 5. Adherence to treatment (remember, do when bad or better, satisfaction) 6. Acceptance of loss (new capability, loss, change, move on).
How You Live	PROM	1. Social determinants of health (education, in control, where I live, money) 2. Social contact (talking, confidants, help, involvement) 3. Loneliness (no one to talk to, left out, alone, lonely) 4. Neighbour relationships (know, trust, share and help) 5. Personal safety (feel safe, respected at home and outside) 6. Digital confidence (own use, peer use, access to help, confidence) 7. Readiness (usage, confidence, like new ideas, keep up to date).
Experience of Care	PREM	1. Patient experience (kind, listen/explain, prompt, organised) 2. Result satisfaction (consultation, treatment, help, next steps) 3. Service integration (staff talk, knowledge, repeat story, collaborate) 4. Data privacy (safe, shared as needed, can check, satisfied) 5. Product confidence (frequent user, confident user, know benefits and problems) 6. User satisfaction (helps me, easy to use, get help, pleased) 7. Training (reaction, learning, behaviour, use results).

Each measure usually has four dimensions (shown in parenthesis) and four options.
PREM, patient-reported experience measures; PROM, patient-reported outcome measures.

different scales and endpoints: Oxford Hip Score: 0 to 48, EQ-5D-3L Index: −1.58 to 1.00, EQ-VAS (Visual Analogue Scale): 0 to 100.[35]

Good reporting practice is to:[2]

► Distinguish between cohort mean scores and individual scores.
► Use common scale endpoints such as 0 for bad and 100 for excellent.
► Avoid using more than two significant figures without good reason—say 67 not 66.7.
► Choose different items to measure different things. This can be tested using statistical tools such as Cronbach's alpha and factor analysis.
► It may be easier to ask everyone than to pick out a representative sample. Statistical significance testing depends on the square root of the sample size.[36]

Range of measures

The range of useful patient-reported measures is broad. PROMs cover more than health status; PREMs cover more than satisfaction with staff. For example, the range of measures developed and used at R-Outcomes Ltd is shown in table 1.[37 38]

Published measures have usually been refined over several years before publication. Always make sure that you know what measures are already available before building your own questionnaire from scratch. Have your own checklist. We value brevity, readability and actionable reports.

CONCLUSIONS

The aims of healthcare can be summed up by the Quintuple Aim[39]:

1. *Health outcomes*—direct impact on patients, families and populations.
2. *Patient experience*—patients' satisfaction with the service they receive.
3. *Cost of care* and unwarranted variation in costs.
4. *Staff satisfaction* to ensure staff well-being and safety.
5. *Health equity*—provide a uniformly excellent service for all.

PROMs and PREMs measure the success of any service against four of the five aims (they do not cover costs). This paper has set out to answer why PROMs and PREMs are not used everywhere. Is it that good measures have not been implemented effectively, or that the measures selected are not fit for purpose, or both?

Innovations, however good, are of no value unless implemented and used. Multiple barriers face the implementation of PROMs and PREMs in routine healthcare. These include the inherent complexity of the healthcare system where, unless well led, many groups and people can prevent or inhibit adoption; also the problems of noise, data sharing (interoperability) and poor planning to ensure that everyone knows and agrees who is to do what, when, where, how and why, and high level leadership. These issues can be recognised in advance and

mitigated or avoided. However, if the measures used are not fit for purpose, either in design or the range of issues addressed, then implementation effort is wasted.

If the measures proposed are not fit for purpose, either in design or the range of issues addressed, then implementation effort is likely to be wasted. However, well-chosen PROMs and PREMs, implemented with care, can help healthcare providers deliver high quality patient outcomes and experience.

ORCID iD
Tim Benson http://orcid.org/0000-0002-2101-1353

REFERENCES

1 Larsson S, Clawson J, Kellar J. *The patient priority: solve health care's value crisis by measuring and delivering outcomes that matter to patients*. New York: McGraw Hill, 2023.
2 Benson T. *Patient-reported outcomes and experience: measuring what we want from PROMs and PREMs*. Cham: Springer, 2022.
3 Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013;346:bmj.f167.
4 Calvert M, Kyte D, Price G, *et al*. Maximising the impact of patient-reported outcome assessment for patients and society. *BMJ* 2019;364:k5267.
5 Bull C, Byrnes J, Hettiarachchi R, *et al*. A systematic review of the validity and reliability of patient-reported experience measures. *Health Serv Res* 2019;54:1023–35.
6 Churruca K, Pomare C, Ellis LA, *et al*. Patient-reported outcome measures (PROMs): a review of generic and condition-specific measures and a discussion of trends and issues. *Health Expect* 2021;24:1015–24.
7 Morris J, Schlepper L, Dayan M, *et al*. Public satisfaction with the NHS and social care in 2022. *The Nuffield Trust* 2023.
8 O'Dowd A. GP patient survey: getting an appointment is harder but decline in satisfaction slows. *BMJ* 2023;382:1629.
9 England NHS. Friends and family test data. Available: https://www.england.nhs.uk/fft/friends-and-family-test-data/ [Accessed 05 Oct 2023].
10 Damschroder LJ. Clarity out of chaos: use of theory in implementation research. *Psychiatry Res* 2020;283:112461.
11 Hodge M. History of the TDS medical information system. In: Blum B, Duncan K, eds. *A history of medical informatics*. ACM Press, 1990: 328–56.
12 Greenhalgh T, Robert G, Macfarlane F, *et al*. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Quarterly* 2004;82:581–629.
13 Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. conceptual framework and item selection. *Med Care* 1992;30:473–83.
14 Brooks R. EuroQol: the current state of play. *Health Policy* 1996;37:53–72.
15 Giordano LA, Elliott MN, Goldstein E, *et al*. Development, implementation, and public reporting of the HCAHPS survey. *Med Care Res Rev* 2010;67:27–37.

16 Damschroder LJ, Aron DC, Keith RE, *et al*. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci* 2009;4:50.
17 Damschroder LJ, Reardon CM, Widerquist MAO, *et al*. The updated consolidated framework for implementation research based on user feedback. *Implement Sci* 2022;17:75.
18 Damschroder LJ, Reardon CM, Opra Widerquist MA, *et al*. Conceptualizing outcomes for use with the consolidated framework for implementation research (CFIR): the CFIR outcomes addendum. *Implement Sci* 2022;17:7.
19 Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health* 1999;89:1322–7.
20 Proctor E, Silmere H, Raghavan R, *et al*. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Ment Health* 2011;38:65–76.
21 Greenhalgh T, Wherton J, Papoutsi C, *et al*. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *J Med Internet Res* 2017;19:e367.
22 Rogers EM. Diffusion of innovations. 5th edn. New York: Free Press, 2003.
23 Miech EJ, Rattray NA, Flanagan ME, *et al*. Inside help: an integrative review of champions in healthcare-related implementation. *SAGE Open Med* 2018;6:2050312118773261.
24 Greenhalgh T, Papoutsi C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. *BMC Med* 2018;16:95.
25 Greenhalgh T, Abimbola S, *et al*. The NASSS framework – a synthesis of multiple theories of technology implementation. In: Scott P, Georgiou A, De Keizer N, eds. *Applied interdisciplinary theory in health informatics*. IOS Press, 2019: 193–204.
26 Langley G, Moen R, Nolan K, *et al*. *The improvement guide: a practical approach to enhancing organizational performance*. San Francisco: Jossey-Bass, 2009.
27 Meyer VM, Benjamens S, Moumni ME, *et al*. Global overview of response rates in patient and health care professional surveys in surgery: a systematic review. *Ann Surg* 2022;275:e75–81.
28 Kahneman D, Sibony O, SunsteinCR. *Noise: a flaw in human judgment*. William Collins: London, 2021.
29 Mullins CF, Coughlan JJ. Noise in medical decision making: a silent epidemic *Postgrad Med J* 2023;99:96–100.
30 Simms LJ, Zelazny K, Williams TF, *et al*. Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychol Assess* 2019;31:557–66.
31 Dillman D, Smyth J, Christian L. *Internet, phone, mail, and mixed-mode surveys: the tailored design method*. John Wiley & Sons, 2014.
32 Basch E. Patient-reported outcomes – harnessing patients' voices to improve clinical care. *N Engl J Med* 2017;376:105–8.
33 Benson T, Grieve G. Principles of health interoperability. In: *Principles of health interoperability: FHIR, HL7 and SNOMED CT*. 4th edn. Cham: Springer, 2021.
34 New collaboration agreement between SNOMED international and LOINC from Regenstieff. SNOMED international. 2022. Available: https://www.snomed.org/news/new-collaboration-agreement-between-snomed-international-and-loinc%C2%AE-from-regenstrief
35 Benson T, Williams DH, Potts HWW. Performance of EQ-5D, *howRu* and Oxford hip & knee scores in assessing the outcome of hip and knee replacements. *BMC Health Serv Res* 2016;16:512.
36 Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use*. 5th edn. Oxford University Press, 2015.
37 Benson T. Measure what we want: a taxonomy of short generic person-reported outcome and experience measures (PROMs and PREMs). *BMJ Open Qual* 2020;9:e000789.
38 Benson T, Benson A. Routine measurement of patient experience. *BMJ Open Qual* 2023;12:e002073.
39 Nundy S, Cooper LA, Mate KS. The quintuple aim for health care improvement: a new imperative to advance health equity. *JAMA* 2022;327:521–2.