

# How do aggregated patient-reported outcome measures data stimulate health care improvement? A realist synthesis

Joanne Greenhalgh<sup>1</sup>, Sonia Dalkin<sup>2</sup>, Elizabeth Gibbons<sup>3</sup>,  
Judy Wright<sup>4</sup>, Jose Maria Valderas<sup>5</sup>, David Meads<sup>6</sup> and  
Nick Black<sup>7</sup>



## Abstract

**Objectives:** Internationally, there has been considerable debate about the role of data in supporting quality improvement in health care. Our objective was to understand how, why and in what circumstances the feedback of aggregated patient-reported outcome measures data improved patient care.

**Methods:** We conducted a realist synthesis. We identified three main programme theories underlying the use of patient-reported outcome measures as a quality improvement strategy and expressed them as nine ‘if then’ propositions. We identified international evidence to test these propositions through searches of electronic databases and citation tracking, and supplemented our synthesis with evidence from similar forms of performance data. We synthesized this evidence through comparing the mechanisms and impact of patient-reported outcome measures and other performance data on quality improvement in different contexts.

**Results:** Three programme theories were identified: supporting patient choice, improving accountability and enabling providers to compare their performance with others. Relevant contextual factors were extent of public disclosure, use of financial incentives, perceived credibility of the data and the practicality of the results. Available evidence suggests that patients or their agents rarely use any published performance data when selecting a provider. The perceived motivation behind public reporting is an important determinant of how providers respond. When clinicians perceived that performance indicators were not credible but were incentivized to collect them, gaming or manipulation of data occurred. Outcome data do not provide information on the cause of poor care: providers needed to integrate and interpret patient-reported outcome measures and other outcome data in the context of other data. Lack of timeliness of performance data constrains their impact.

**Conclusions:** Although there is only limited research evidence to support some widely held theories of how aggregated patient-reported outcome measures data stimulate quality improvement, several lessons emerge from interventions sharing the same programme theories to help guide the increasing use of these measures.

## Keywords

patient-reported outcome measures, quality improvement, realist synthesis

## Introduction

Internationally, there has been considerable debate about the role of data in supporting quality improvement in health care.<sup>1</sup> The most recent metric to be mobilized for this purpose are patient-reported outcome

<sup>1</sup>Associate Professor, School of Sociology and Social Policy, University of Leeds, UK

<sup>2</sup>Senior Lecturer, Department of Social Work, Education and Community Well Being, Northumbria University, Newcastle, UK

<sup>3</sup>Senior Research Scientist, Nuffield Department of Population Health, University of Oxford, UK

<sup>4</sup>Senior Information Specialist, Leeds Institute of Health Sciences, University of Leeds, UK

<sup>5</sup>Professor of Health Services and Policy Research, University of Exeter Medical School, UK

<sup>6</sup>Associate Professor of Health Economics, Leeds Institute of Health Sciences, University of Leeds, UK

<sup>7</sup>Professor of Health Services Research, London School of Hygiene and Tropical Medicine, UK

### Corresponding author:

Joanne Greenhalgh, School of Sociology and Social Policy, University of Leeds, Leeds LS2 9JT, UK.

Email: j.greenhalgh@leeds.ac.uk

measures (PROMs): questionnaires that measure patients' perceptions of the impact of a condition and its treatment on their health.<sup>2</sup> PROMs aim to capture and quantify outcomes that are important to patients. Some countries have begun routinely collecting aggregated PROMs as an indicator of health care quality for some specialities and procedures, or are moving towards doing so.<sup>3</sup> In the US, PROMs form part of the Health Outcomes Survey, used by the Centres for Medicare and Medicaid Services to evaluate the Medicare Advantage programme.<sup>4</sup> In England, PROMs are routinely collected for all patients undergoing surgery for hip and knee replacement, varicose vein surgery and hernia repair<sup>3</sup> and for patients accessing psychological therapies as part of the Increasing Access to Psychological Therapies initiative.<sup>5</sup> Collecting data for the English PROMs programme, covering about 250,000 patients, is estimated to cost £825,000 annually.<sup>6</sup>

The routine collection of aggregated PROMs data in England was initially driven by economic considerations to improve demand management for high-volume elective procedures. As the programme matured, it highlighted questions about the timing of surgery in relation to patient need and provided opportunities to use these data to compare the effectiveness of different brands of implants and different types of surgical techniques. More recently, the focus has shifted on the use of PROMs data in managing the performance of providers (organizations and individuals) and supporting patients' choice of hospital.<sup>2</sup>

The use of aggregated PROMs data as a tool for quality assurance and improvement is relatively recent, and there is inevitably only limited evidence about their use for these purposes. However, we can draw on evidence about the use of other forms of performance data which shares similar ideas and assumptions about how they are intended to work.<sup>7</sup> Learning from these programmes across different countries can inform the implementation of PROMs feedback. It has been hypothesized that such data can improve the quality and outcomes of patient care either through a 'selection' pathway (whereby patients choose a high-quality hospital) or through a 'change' pathway (whereby providers initiate quality improvement).<sup>8</sup> Provider initiatives include those seeking to preserve their reputation by initiating top-down managerial interventions,<sup>9</sup> changing social norms regarding what constitutes organizational effectiveness or motivating clinicians to become involved in quality improvement.<sup>10</sup>

Previous reviews of quantitative evidence suggest there is little support for the 'selection' pathway but there is evidence to support the 'change' pathway.<sup>11,12</sup> These reviews show that changes made in response to performance data vary in their impact on provider behaviour and health outcomes, while other work has

noted the importance of contextual features in shaping the impact of such interventions.<sup>13</sup> However, there has been no systematic attempt to synthesize how context shapes the processes or *mechanisms* through which the public or private feedback of performance data (such as PROMs) changes provider behaviour and improves patient care and patient outcomes.

We carried out a realist synthesis to examine how and in what circumstances the feedback and public reporting of aggregated PROMs and other forms of performance data improve the quality of patient care. Specifically, we sought to (1) identify and catalogue the different ideas and assumptions or *programme theories* underlying how the feedback of aggregated PROMs data is intended to improve patient care; and (2) review the international evidence to examine how contextual factors shape the mechanisms through which the feedback of PROMs and other performance data leads to intended and unintended changes in provider behaviours and patient outcomes.

## Methods

### Rationale

Realist synthesis enables an understanding of how, why and in what circumstances the feedback of PROMs data improves care.<sup>14</sup> It involves a process of identifying, testing and refining theories about how, when and why the intervention leads to both intended and unintended outcomes. These are expressed as context–mechanism–outcome (C–M–O) configurations and they take the form of 'if-then' propositions which hypothesize 'in these circumstances (*context*), stakeholders respond by changing their reasoning and behaviour in this way (*mechanism*), which leads to this pattern of outcomes'. The unit of analysis of realist synthesis is not the intervention but the programme theories that underpin it. Our scoping searches indicated a paucity of evidence on the use and impact of aggregated PROMs data on patient care. Therefore, evidence from other forms of performance data feedback and public reporting was incorporated if it shared the same programme theories. In these circumstances, realist synthesis is a valuable review methodology; its focus is on reviewing evaluations of interventions which share the same programme theories, even if these ideas take the form of different interventions. The protocol was published<sup>15</sup> and the RAMESES guidelines<sup>16</sup> have been followed in reporting our findings.

### Identifying programme theories

Programme theories underlying PROMs feedback were identified through an analysis of policy documents,

'position' pieces or reviews, comments, letters and editorials (search strategy in online Appendix 1), as well as informal meetings and a two hour workshop with clinicians, managers, policymakers and patients to verify, expand and refine the theories. We also looked for similarities between the ideas and assumptions articulated in policy documents regarding how PROMs were intended to work and substantive theories at a higher level of abstraction, such as theories of audit and feedback,<sup>17</sup> benchmarking<sup>18</sup> and public disclosure of performance data.<sup>14</sup> This enabled us to (1) to identify interventions that shared the same programme theories and thus could form the evidence base for our synthesis (especially where the evidence base for PROMs was sparse), and (2) to develop a series of 'if-then' propositions that formed the basis of theory testing against empirical studies in our evidence synthesis.

### *Searching for empirical evidence to test the programme theories*

Our search strategies to identify studies that explored providers' opinions, their experiences of using and responding to PROMs and other performance data, and the impact of PROMs and other performance data on quality improvement were carried out in October 2014 and are shown in online Appendix 1. Other performance data included mortality report cards and patient experience measures. We searched Embase Classic+Embase (Ovid), HMC Health Management Information Consortium (Ovid), (Ovid) MEDLINE<sup>®</sup> and (Ovid) MEDLINE<sup>®</sup> In-Process & Other Non-Indexed Citations to October 2014. We also carried out backwards citation tracking of a systematic review<sup>11</sup> and five papers<sup>19–23</sup> that were purposively selected to carry out a 'pilot' synthesis. Further, we conducted an additional electronic search of (Ovid) MEDLINE<sup>®</sup> and Embase Classic+Embase (Ovid) focusing on patient experience, which was supplemented by papers from the project team's personal libraries.

We developed a set of inclusion and exclusion criteria (online Appendix 2) to identify papers relevant to testing our programme theories. Following several iterative screening procedures, we included an initial set of 51 papers. As the synthesis progressed and we refined our theories, we excluded 23 papers that no longer contributed to our emerging synthesis. Supplementary searches and citation tracking of key papers remaining in the synthesis yielded an additional 30 papers, so that we included a total of 58 papers in the synthesis. A subsequent search in October 2016 identified two additional systematic reviews and a further three studies (five papers in total) that contributed to theory testing and refinement (online Appendix 1),

yielding a final total of 63 papers the we included in the review (Figure 1).

### *Data extraction, quality assessment and synthesis*

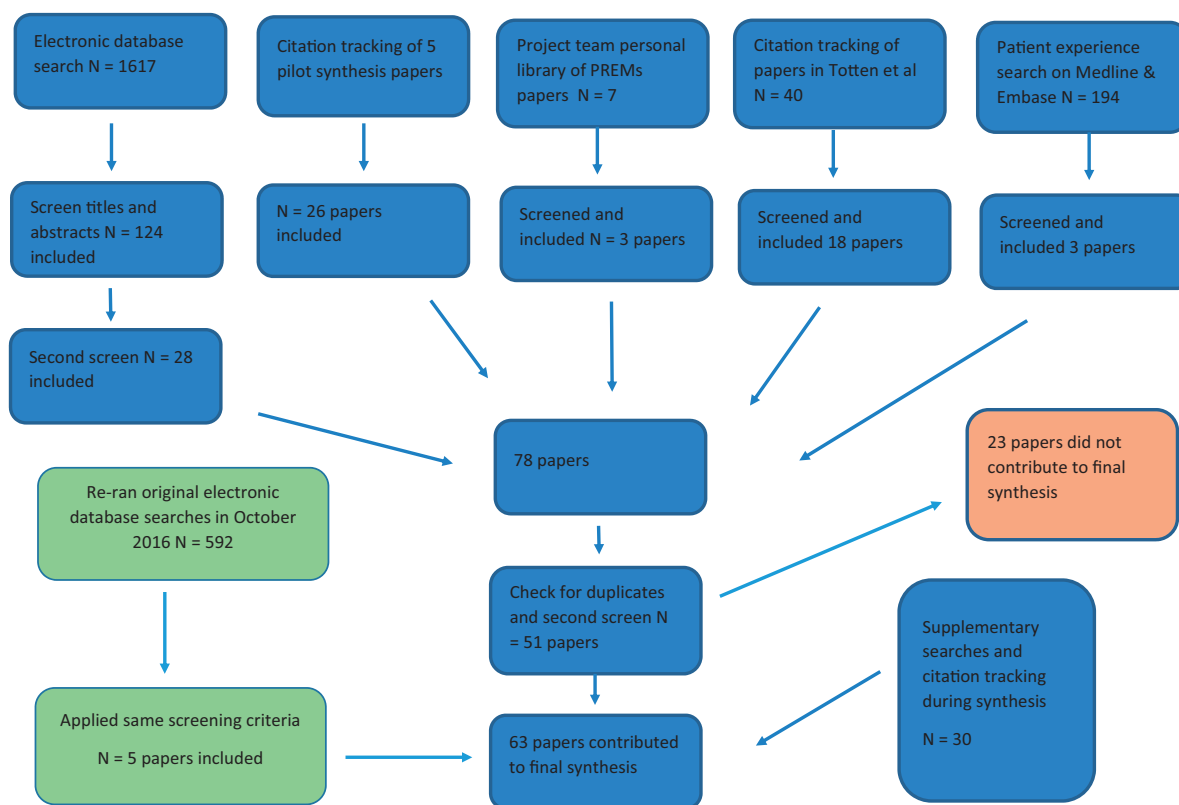
Data extraction, quality assessment, additional searching and synthesis were intertwined and informed each other. We developed a data extraction template to record details of study title, aims, methodology and quality assessment, main findings (including authors' interpretations) and links to programme and abstract theory. Quality appraisal was done on a case-by-case basis, as appropriate to the method utilized in the original study. Each fragment of evidence was appraised, as it was extracted, for its relevance to theory testing and the rigour with which it has been produced.<sup>24</sup> In many instances, only a subset of findings from each study that related specifically to the theory being tested was included in the synthesis. Therefore, quality appraisal related specifically to the validity of the causal claims made in these findings.

We conducted a 'pilot' synthesis on a purposively selected sample of five papers<sup>19–23</sup> to initially test our theories and then conducted the synthesis across all papers to further test, refine and consolidate these theories. After a first reading of the papers, we produced a table summarizing each study's findings by key contextual factors, mechanisms and outcomes. We used these tables and a rereading of the papers to begin to develop ideas about how these findings might be brought together as C–M–O configurations. This was both a within- and a cross-study analysis. Within studies, we identified patterns in which particular clusters of contextual factors gave rise to a particular provider response or responses. We compared these patterns across studies and attempted to explain why similar or different patterns might have arisen. We held regular team meetings to discuss and refine our synthesis.

## **Results**

### *Programme theories: how is the feedback of aggregated PROMs data expected to improve patient care?*

We identified three main programme theories underlying how aggregated PROMs and performance are intended to improve patient care (online Table 1): supporting patient choice<sup>25,26</sup>; making providers accountable to commissioners, regulators and the public<sup>5,25</sup> and enabling providers to compare their performance with others (often referred to as benchmarking).<sup>25</sup> We also identified a number of contextual factors that shape the mechanisms through which the feedback and public reporting of PROMs and performance data lead to



**Figure 1.** Flow chart of identification and inclusion of empirical studies for theory testing.

intended outcomes (the initiation of quality improvement activities and subsequent improvements in patient care) or unintended outcomes (e.g. dismissing or ignoring data, tunnel vision and gaming) (online Table 1). For the purposes of theory testing and informed by theories of public disclosure,<sup>7</sup> of audit and feedback<sup>17</sup> and of benchmarking,<sup>18</sup> we grouped these as follows: The degree to which data are publicly rather than only privately disclosed, the use of financial incentives and sanctions, the perceived credibility of these data and the ‘actionability’ of data (the extent to which performance data identify practical steps that could be implemented to improve services).

To test and refine our programme theories and understand how contextual factors shaped the mechanisms through which the feedback of PROMs data improved patient care, we developed a series of nine ‘if-then’ propositions, with a number of linked sub-propositions (online Table 2). Research evidence regarding each of the nine propositions is presented under each of the three main theories (online Table 2).

### Patient choice theory

This theory proposes that PROMs data might improve services through supporting patient choice as providers would respond by improving care in

order to protect their market share. We tested and refined this theory through Propositions 3, 4, 5 and 6 (online Table 2). In general, providers did not see a change in their market share following the public reporting of performance data (Proposition 3).<sup>11,27</sup> Nonetheless, one US study suggested they were more likely to experience a change in market share if performance data provided ‘new information’ and revealed that previously ‘good’ hospitals had performed poorly, i.e. if performance data damaged their reputation (Proposition 4).<sup>28</sup> One UK study indicated that demand for hospitals was more responsive to changes in PROMs scores than to readmission or mortality rates, but the overall change in market share was very small.<sup>29</sup> Patients were more likely to choose a non-local hospital if they had a bad previous experience with their local hospital.<sup>30</sup> They indicated that cleanliness and a hospital’s reputation were most important in their choice.

Patients did not use formal, publicly reported information about service quality to inform their choice of hospital (Proposition 6) but instead relied on their previous personal experience of the provider, the opinions of friends and family, and advice from their primary care doctor<sup>30</sup> who also based referral decisions on their relationships and past experiences with providers.<sup>30,31</sup>



Hospital providers were more concerned about the consequences of public reporting on their reputation than on their market share, especially if they were a classified as a poor performer (Proposition 4).<sup>9,32</sup> However, they did perceive that damage to their reputation could have an impact on whether patients chose their hospital,<sup>19,33,34</sup> which they felt was also influenced by adverse media reporting of publicly reported data.<sup>19</sup> Providers were also motivated to improve the quality of care they offered because they wished to be seen to be as good as or better than their peers (Proposition 5).<sup>32,35</sup> Furthermore, taking steps to improve the quality of care was perceived as enhancing a provider's reputation, which in turn meant that patients were more likely to be referred to them.<sup>22</sup>

### Accountability theory

This theory suggests that PROMs data will enable commissioners and regulators to hold providers to account and, as a result, providers will improve because they fear the imposition of sanctions. We tested this theory through propositions 2, 7, 8 and 9. There is evidence to indicate that public reporting places additional pressure on providers to respond, particularly poor performers (Proposition 2).<sup>9,32</sup> When data were fed back privately, with no public reporting or financial incentives, providers either ignored them or focused on altering their data collection practices.<sup>21,36</sup> There is some evidence to suggest that greater improvements in the quality of patient care occurred when providers are subjected to both financial incentives and public reporting than when they are subjected to either initiative alone (Proposition 7).<sup>37</sup> However, financial incentives only led to a short-term improvement in quality if they focused on activities that providers already performed well.<sup>38</sup> In addition, the extent of any improvement was limited because providers would stop attempts to improve once they reached the threshold at which they would receive the maximum amount of remuneration.<sup>38</sup>

The evidence also suggests that providers perceived mandatory public reporting initiated by regulators or government as politically motivated and focused on aspects of care that mattered to those bodies rather than to clinicians (Proposition 8).<sup>19,22,34,39</sup> Financial incentives led to 'tunnel vision' or effort substitution.<sup>38</sup> Providers subjected to financial incentives based on public reporting of indicators that they did not feel were valid or clinically relevant led to manipulation or gaming of the data.<sup>19,40,41</sup> This was not always the result of deliberate attempts to 'cheat' the system on the part of providers; rather, the use of financial rewards created perverse incentives, which were at odds with the inherent clinical complexity of managing long-term

conditions such as depression.<sup>40,41</sup> For example, in primary care in England, if patients diagnosed with depression did not return for a follow-up appointment, it was not possible for the GP to use a PROM on this occasion, with the consequence that the practice would lose income under the Quality Outcome Framework (pay-for-performance scheme in primary care). Under these conditions, clinicians managed the uncertainty of diagnosis by raising their thresholds for diagnosing depression in order to ensure they were not financially penalized.<sup>40,41</sup>

While mandatory public reporting may help focus the attention of hospital leaders and clinicians on quality,<sup>20,39</sup> providers were more likely to respond to nationally produced public reports when they were consistent with their internally collected data.<sup>22</sup> This was because they trusted the latter data, which provided more detailed information and insight.<sup>22</sup> Mandated publicly reported data, particularly those focusing on outcomes, often did not enable providers to identify the causes of poor care (Proposition 9). Thus, additional investigations were needed, which required resources and staff with the 'know-how' and capacity to carry them out.<sup>22</sup>

We tested the hypothesis that data on patients' experience of care (rather than outcomes) were more likely to lead to improvement as they provided a clearer indication of which care processes needed to be improved, could be fed back in a timely manner and could be reported at hospital ward as well as institutional level (Proposition 9). Providers held more favourable attitudes towards process of care and patient experience measures than outcome and cost data.<sup>34</sup> Although patient experience data identified potential problems, these were often already known to staff<sup>42</sup> and this did not always lead to improvements.<sup>43</sup> Where changes were made, these tended to focus on the organizational or functional aspects of care (sometimes referred to as 'easy stuff'), rather than the more challenging relational aspects such as the behaviour and communication skills of clinicians.<sup>42,43</sup> Interventions were more likely to succeed if they did not require complex organizational changes or changes in clinical behaviour.<sup>44</sup> As with outcome data, attempts to respond to issues highlighted by patient experience data could create new problems; for example, Davies et al.<sup>44</sup> found communication skills training for clinical staff led to a slight increase in patients reporting that doctors explained things to them in a way that was easy to understand but also a decline in the percentage of patients reporting that doctors spent enough time with them. These findings suggest that significant and sustained improvement is only achievable with system- and organization-wide strategies.<sup>44</sup>

### Provider benchmarking theory

The third theory hypothesized that PROMs will enable providers to compare their own performance with those of their peers and thus seek to improve their performance. We tested and refined this theory through propositions 2, 5 and 8. The evidence suggested that mechanisms such as clinicians' intrinsic motivation to improve, aspirations to be as good as, or better than, their peers, or a desire to learn from the best practices of others were more likely to be triggered when the public reporting of data was clinically led (Proposition 5 and 8).<sup>35,45</sup> Clinician involvement led to the selection of clinically relevant indicators and, in turn, clinical ownership of the sources of data, indicator specification and methods of case-mix adjustment, making it difficult for them to dismiss or ignore the data.<sup>34,35</sup> Under these circumstances, providers took steps to improve patient care to be as good as or better than their peers, and did so through sharing and learning from best practices.<sup>35</sup> However, even under these conditions, action to improve patient care depended on the provider's skills in quality improvement; those with more experience in quality improvement were more likely to make sustainable improvements.<sup>45</sup>

### Discussion

We identified eight lessons about the use of performance data for quality improvement which could be applied in the future implementation of aggregated PROMs data internationally.

First, as regards patient choice, patients or their agents (e.g. GPs) rarely use any published performance data when selecting a provider. This insight is not new and has been identified by previous systematic reviews.<sup>11,12</sup> Our review found that patients prefer to rely on their own past experiences and those of friends and family when choosing a hospital. Some have speculated that patients may be more likely to use PROMs to select a hospital because they would find these data more relevant.<sup>26</sup> Although one study suggested that demand is more responsive to variation in hospital performance measured by PROMs than mortality and readmission rates,<sup>29</sup> it was not clear that this demand was stimulated by patients or their agents consulting quality data.

Second, when clinicians perceived that performance indicators were not credible but were incentivized to collect them, gaming or manipulation of data occurred. There is some evidence that clinicians remain sceptical about the validity of using patients' subjective assessments of outcome as an indicator of the quality of patient care.<sup>21,33</sup>

Third, the perceived motivation behind public reporting is an important determinant of how providers

respond. In particular, public reporting programmes that are initiated by government or insurers were often perceived as being driven by political motives rather than a desire to improve patient care and could risk leading to 'tunnel vision'.<sup>19</sup> Organizations with less experience of quality improvement initiatives were less likely to make sustainable improvements in response to performance data.<sup>45</sup>

Fourth, the lack of timeliness of performance data acted as a constraint to providers' use in quality improvement. In the case of PROMs data, this is a significant problem, because of the time required to collect PROMs data, link them to hospital administrative data, adjust for case mix and produce indicators. One option is to collect PROMs data electronically, but this brings its own challenges in terms of response rates and engagement from key stakeholders.<sup>46</sup>

Fifth, performance data based on outcomes do not provide information on the cause of poor care, which in turn constrains their use in quality improvement activities. As such, performance data focused on outcomes act as 'tin openers' not 'dials'.<sup>47</sup> In the case of the PROMs data, providers are classified as 'statistical' outliers but this does not necessarily mean that they are providing poor care and nor does it inform them about the possible causes for any poor care. Providers are expected to conduct further investigations to identify alternative explanations for their outlier status and, if none are found, explore the possible cause(s) of their outlier status.

This leads to the sixth lesson. Providers could only conduct these investigations if they routinely collected their own quality data and had the necessary infrastructure in place to embark on further investigations. Providers who were more experienced in quality improvement activities were more likely to have these structures in place, although the direction of causation is not clear.<sup>22,45</sup> Therefore, providers classed as 'outliers' need more support for and guidance on how to collect their own data, how to interpret the data and how to explore the possible causes of observed variation.

Seventh, even when providers were given support to carry out additional investigations, implementing changes remained a challenge.<sup>44</sup> This suggests that bringing about significant, sustainable quality improvement requires a system-wide approach to change. The investigation of key factors involved in this process has been explored in other reviews.<sup>48</sup>

Eighth, PROMs data are only one of many pieces of performance data that providers have to respond to and when providers are not classed as an outlier, PROMs data run the risk of being ignored as providers tend to focus on other areas of poor performance. Greater improvements in patient care could

be realized if all providers used PROMs data to improve patient care, not just those who were negative outliers. To support this and avoid tunnel vision, further support and guidance are needed to enable providers to integrate and interpret PROMs data in the context of other performance data. For hospital providers, this will require not only effective co-ordination across different hospital departments but also efficient lines of communication between the hospital board and clinical staff. Dissemination should involve top-down and bottom-up approaches so that clinical staff are aware of and know how to interpret performance data but also so that staff can bring their insights about the possible causes of problems to hospital leaders.<sup>1</sup>

### Strengths and limitations

This review is the first attempt to systematically synthesize the international literature to explore how different contextual factors, that is, the ways in which aggregated PROMs and other performance data are implemented and the systems into which they are introduced, shape the mechanisms through which the feedback of performance data do or do not achieve their intended objective of stimulating quality improvement and, in turn, improved patient care.

Our review has a number of limitations. A significant challenge when conducting a realist synthesis is drawing boundaries around the review.<sup>14</sup> Our initial decisions regarding the range of substantive theories to draw upon and thus the formulation of our propositions for theory testing meant that some theoretical perspectives and literatures were inevitably omitted from our review. Using Harrison's and Ahmad's<sup>49</sup> work on scientific bureaucratic medicine and Foucault's<sup>50</sup> ideas of disciplinary power and governmentality would have provided a useful alternative lens to test of our accountability theory and enabled us to explore how and when performance data function as curb to clinical autonomy<sup>23</sup> and/or act as a form of self-surveillance.<sup>51</sup>

We focused on understanding how providers responded to performance data in terms of the initiation (or not) of quality improvement activities. Although testing these theories led us to consider how some characteristics of organizations supported or limited the success of the implementation of quality improvement activities, we did not comprehensively review this literature.<sup>52</sup> Similarly, there are other explanations for why patients may not utilize quality data that we did not explore.<sup>53</sup> We also acknowledge that patient experience data enable providers to understand what matters to patients more generally which may inform training, rather than, or in addition to, quality improvement activities.<sup>54</sup>

Literature on the use of aggregated PROMs data is limited as the use of these measures for quality assurance and improvement is still in its infancy. Supplementing this literature with evidence from similar types of performance data helped provide a fuller picture but some caution is required in extrapolating from feedback for these other types of data.

### Conclusions

There has been considerable interest internationally about the role of PROMs feedback in stimulating quality improvement initiatives.<sup>3</sup> The Organisation for Economic Co-operation and Development has recently considered the potential use of PROMs to facilitate improvement in countrywide health systems and is planning to incorporate them into its measurement and performance framework.<sup>55</sup> The English National PROMs programme has recently been reviewed and some have called for a shift in purpose.<sup>56</sup> Our review synthesized evidence from interventions which share the same programme theories to understand how context shapes the mechanisms through which PROMs feedback may support quality improvement activities. The lessons drawn from our review can be used internationally to inform the future implementation of programmes which feedback and publicly report aggregated PROMs data.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was funded by NIHR Health Services and Delivery Research 12/136/31. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. EG was also funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care Oxford at Oxford Health NHS Foundation Trust.

### References

1. Martin GP, McKee L and Dixon-Woods M. Beyond metrics? Utilizing 'soft intelligence' for healthcare quality and safety. *Soc Sci Med* 2015; 142: 19–26.
2. Valderas JM and Alonso J. Patient reported outcome measures: a model-based classification system for research and clinical practice. *Qual Life Res* 2008; 17: 1125–1135.
3. Van der Wees PJ, Nijhuis-Van der Sanden MWG, Ayanian JZ, et al. Integrating the use of patient-reported outcomes for both clinical practice and performance

- measurement: views of experts from 3 countries. *Milbank Quart* 2014; 92: 754–775.
4. Centres for Medicare and Medicaid Services (CMS). *Health Outcomes Survey (HOS)*. Baltimore, MD: CMS, 2016.
  5. National IAPT Programme Team. *The IAPT data handbook: guidance on recording and monitoring outcomes to support local evidence-based practice*. London: Crown Copyright, 2011.
  6. England N. *National patient reported outcome measures (PROMs) programme consultation*. NHS England, 2016, available at: <https://www.engage.england.nhs.uk/consultation/proms-programme/> retrieved 30 October 2017).
  7. Pawson R. Evidence and policy and naming and shaming. *Policy Stud* 2002; 23: 211–230.
  8. Berwick DM, James B and Joel Coye M. Connections between quality measurement and improvement. *Med Care* 2003; 41: 1–30.
  9. Hibbard J, Stockard J and Tusler M. Hospital performance reports: impact on quality, market share, and reputation. *Health Aff* 2005; 24: 1150–1160.
  10. Contandriopoulos D, Champagne F and Denis JL. The multiple causal pathways between performance measures' use and effects. *Med Care Res Rev* 2014; 71: 3–20.
  11. Totten AM, Wagner J, Tiwari A, et al. Closing the quality gap: revisiting the state of the science (vol. 5: public reporting as a quality improvement strategy). *Evid Rep/ Technol Assess* 2012; 208(5): 1–645.
  12. Fung CH, Lim Y-W, Mattke S, et al. Systematic review: the evidence that publishing patient care performance data improves quality of care. *Ann Intern Med* 2008; 148: 111–123.
  13. Lemire M, Demers-Payette O and Jefferson-Falardeau J. Dissemination of performance information and continuous improvement: a narrative systematic review. *J Health Organ Manag* 2013; 27: 449–478.
  14. Pawson R. *Evidence based policy: a realist perspective*. London: Sage, 2006.
  15. Greenhalgh J, Pawson R, Wright J, et al. Functionality and feedback: a protocol for a realist synthesis of the collation, interpretation and utilisation of PROMs data to improve patient care. *BMJ Open* 2014; 4: e005601.
  16. Wong CK, Greenhalgh T, Westhorp G, et al. RAMESES publication standards: realist synthesis. *BMC Med* 2013; 11: 21.
  17. Kluger AN and DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol Bull* 1996; 119: 254–284.
  18. van Helden GV and Tilemma S. In search of a benchmarking theory for the public sector. *Financial Accountability Manag* 2005; 21: 337–361.
  19. Mannion R, Davies H and Marshall M. Impact of star performance ratings in English acute hospital trusts. *J Health Serv Res Policy* 2005; 10: 18–24.
  20. Hafner JM, Williams SC, Koss RG, et al. The perceived impact of public reporting hospital performance data: interviews with hospital staff. *Int J Qual Health Care* 2011; 23: 697–704.
  21. Boyce M, Browne J and Greenhalgh J. Surgeon's experiences of receiving peer benchmarked feedback using patient-reported outcome measures: a qualitative study. *Implement Sci* 2014; 9: 84.
  22. Davies H. Public release of performance data and quality improvement: internal responses to external data by health care providers. *Qual Health Care* 2001; 10: 104–110.
  23. Exworthy M, Wilkinson E, McColl A, et al. The role of performance indicators in changing the autonomy of the general practice profession in the UK. *Soc Sci Med* 2003; 56: 1493–1504.
  24. Pawson R. Digging for nuggets: how bad research can yield good evidence. *Int J Soc Res Methodol* 2006; 9: 127–142.
  25. Health and Social Care Information Centre. *Monthly patient reported outcome measures (PROMs) in England*. London: Health and Social Care Information Centre, 2015.
  26. Marshall M and McLoughlin V. How do patients use information on health providers? *BMJ* 2010; 341: c5272.
  27. Chassin MR. Achieving and sustaining improved quality: lessons from New York State and cardiac surgery. *Health Aff* 2002; 21: 40–51.
  28. Dranove D and Sfekas A. Start spreading the news: a structural estimate of the effects of New York hospital report cards. *J Health Econ* 2008; 27: 1201–1207.
  29. Gutacker N, Siciliani L, Moscelli G, et al. Choice of hospital: which type of quality matters? *J Health Econ* 2016; 50: 230–246.
  30. Dixon A, Robertson R, Appleby J, et al. *Patient choice: how patients choose and how providers respond*. London: The King's Fund, 2010.
  31. Greener I and Mannion R. Patient choice in the NHS: what is the effect of patient choice policies on patients and relationships in health economies. *Public Money Manag* 2009; 29: 95–100.
  32. Hibbard J, Stockard J and Tusler M. Does publicizing hospital performance stimulate quality improvement efforts? *Health Aff* 2003; 24: 84–94.
  33. Hildon Z, Allwood D and Black N. Patients' and clinicians' views of comparing the performance of providers of surgery: a qualitative study. *Health Expect* 2012; 8(3): 366–378.
  34. Mehrotra A, Bodenheimer T and Dudley RA. Employers' efforts to measure and improve hospital quality: determinants of success. *Health Aff* 2003; 22: 60–71.
  35. Greer AL. *Embracing accountability: physician leadership, public reporting, and teamwork in the Wisconsin Collaborative for Healthcare Quality*. New York: Commonwealth Fund, 2008.
  36. Mannion R and Goddard M. Public disclosure of comparative clinical performance data: lessons from the Scottish experience. *J Eval Clin Pract* 2003; 9: 277–286.
  37. Alexander JA, Maeng D, Casalino LP, et al. Use of care management practices in small-and medium sized-physician groups: do public reporting of physician quality and financial incentives matter? *Health Serv Res* 2013; 48: 376–387.



38. Doran T, Kontopantelis E, Valderas JM, et al. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. *BMJ* 2011; 342: d3590.
39. Pham HH, Coughlan J and O'Malley AS. The impact of quality-reporting programs on hospital operations. *Health Aff* 2006; 25: 1412–1422.
40. Dowrick C, Leydon GM, McBride A, et al. Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. *BMJ* 2009; 338: b663.
41. Mitchell C, Dwyer R, Hagan T, et al. Impact of the QOF and the NICE guideline in the diagnosis and management of depression: a qualitative study. *Br J Gen Pract* 2011; 61: e279–289.
42. Boiko O, Campbell JL, Elmore N, et al. The role of patient experience surveys in quality assurance and improvement: a focus group study in English general practice. *Health Expect* 2015; 18: 1982–1994.
43. Elliott MN, Lehrman WG, Goldstein EH, et al. Hospital survey shows improvements in patient experience. *Health Aff* 2010; 29: 2061–2067.
44. Davies E, Shaller D, Edgman-Levitan S, et al. Evaluating the use of a modified CAHPS survey to support improvements in patient-centred care: lessons from a quality improvement collaborative. *Health Expect* 2008; 11: 160–176.
45. Smith MA, Wright A, Queram C, et al. Public reporting helped drive quality improvement in outpatient diabetes care among Wisconsin physician groups. *Health Aff* 2012; 31: 570–577.
46. Keetharuth A, Mulhern B, Wong R, et al. *Supporting the routine collection of patient reported outcome measures (PROMs) for the National Clinical Audit. Work Package 2. How should PROMs data be collected?* Report to the Department of Health Policy Research Programme. Sheffield: University of Sheffield and University of York, 2015.
47. Carter N, Day P and Klein R. *How organisations measure success: the use of performance indicators in government*. London: Routledge, 1995.
48. Pawson R, Greenhalgh J, Brennan C, et al. Do reviews of healthcare interventions teach us how to improve health-care systems. *Soc Sci Med* 2013; 114: 129–137.
49. Harrison S and Ahmad WIU. Medical autonomy and the UK state 1975 to 2025. *Sociology* 2000; 34: 129–146.
50. Foucault M. The subject and power. *Critical Inquiry* 1982; 8(4): 775–795.
51. Martin GP, Leslie M, Minion J, et al. Between surveillance and subjectification: professionals and the governance of quality and patient safety in English hospitals. *Soc Sci Med* 2013; 99: 80–88.
52. Kaplan HC, Brady PW, Dritz MC, et al. The influence of context on quality improvement success in health care: a systematic review of the literature. *Milbank Quart* 2010; 88: 500–559.
53. Hibbard JH and Peters E. Supporting informed consumer health care decisions: data presentation approaches that facilitate the use of information in choice. *Annu Rev Public Health* 2003; 24: 413–433.
54. Coulter A, Locock L, Ziebland S, et al. Collecting data on patient experience is not enough: they must be used to improve care. *BMJ* 2014; 348: g2225.
55. OECD. *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: OECD, 2017.
56. Browne J, Cano SJ and Smith S. Using patient-reported outcome measures to improve health care time for a new approach. *Med Care* 2017; 55: 901–904.