https://ijhpm.com
Int J Health Policy Manag 2026;15:10444

OPEN ACCESS

doi 10.34172/ijhpm.10444

IJHPM
International Journal of Health Policy and Management

Commentary

Making the Evaluation of Quality Indicators Count in Improving Patient Care

Comment on "Comprehensive Evaluation of Quality Indicators: Analyzing the Dutch Breast Cancer Audit"

Perla J. Marang-van de Mheen[1*], Jakko van Kampen[2], Jop Groeneweg[1,2]

Abstract

Quality indicators are widely used, and periodic evaluation may help to ensure their usefulness in daily practice. Verheul and colleagues provide an approach for a data-driven evaluation to assess the feasibility, discriminative ability, validity, and reliability of 19 indicators used to benchmark hospital performance. They show good feasibility and discriminative ability, but also poor reliability in detecting true hospital differences which is concerning particularly for publicly reported indicators. Using their evaluation as an example, we consider in this commentary what a good evaluation should look like. Four key questions are proposed to guide choices for an evaluation approach. We highlight the importance of distinguishing between benchmarking and improvement purposes, defining the appropriate unit and criteria of evaluation, specifying thresholds for action based on the evaluation results and considering stakeholder perspectives. Aligning the evaluation approach with the intended use of indicators is critical to ensure indicators contribute to improving patient care.

Keywords: Quality Indicators, Evaluation, Benchmarking, Improvement, Stakeholder Perspectives

Copyright: © 2026 The Author(s); Published by Kerman University of Medical Sciences. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Marang-van de Mheen PJ, van Kampen J, Groeneweg J. Making the evaluation of quality indicators count in improving patient care: Comment on "Comprehensive evaluation of quality indicators: analyzing the Dutch Breast Cancer Audit." *Int J Health Policy Manag.* 2026;15:10444. doi:10.34172/ijhpm.10444

Article History:
Received: 13 October 2026
Accepted: 17 November 2026
ePublished: 22 November 2026

***Correspondence to:**
Perla J. Marang-van de Mheen
Email:
p.j.marang-vandemheen@tudelft.nl

Introduction

Should we care about quality indicators if they do not add value for healthcare organisations in improving care for their patients? A recent study showed that Dutch healthcare professionals spent on average 52 minutes per working day collecting data on various quality registrations, with only 36% of the quality indicators perceived as useful to improve the quality of care.[1] Perceived usefulness is important as only when these indicators are used effectively, they will have impact in improving care. Therefore, it is important to periodically examine, evaluate and refine indicator sets to ensure their (continued) usefulness.

The study by Verheul and colleagues is a welcome contribution to the literature as they show how a set of quality indicators can be evaluated systematically based on empirical data, in this case the current set used by the NABON Breast Cancer Audit (NBCA).[2] They evaluated 19 indicators relevant for between-hospital comparisons on 4 criteria (feasibility, discriminative ability, validity, and reliability), using NBCA data on all surgically treated breast cancer patients in the Netherlands between 2021-2023. Of these 19 indicators, 9 were publicly reported and the majority had not been evaluated previously. They found good feasibility and discriminative ability, but only 6 indicators met 3 of the 4 criteria and none met all 4. The authors conclude that the rankability of indicators (assessing reliability) is a concern especially for public reporting to ensure fair hospital comparisons. They recommend reporting indicator scores on multiple years or at an aggregate level for more reliable public reporting, while also acknowledging that this contradicts short cycle feedback which is needed for improvement purposes. When indicators do not meet all criteria, Verheul et al suggest these should not be publicly reported but can still "fuel collective learning and improving."

Where much has been written about criteria for a good indicator, less attention has been paid to what a good evaluation of a set of indicators should look like to contribute to the set of indicators being useful in daily practice. The paper of Verheul et al shows there is potentially much to be gained if we draw lessons from such evaluation. Using their evaluation as an example, we draw attention to four key questions that can be considered when designing the evaluation of an indicator set, highlighting several decision points with their implications.

What Is the Focus of the Evaluation?

The first question is whether the evaluation is focused on the properties of individual indicators within the set or the indicator set as a whole. The study by Verheul et al is an

Full list of authors' affiliations is available at the end of the article.

example of the first type, where all indicators are evaluated independently on the same four criteria. When the focus is instead on the indicator set as a whole, then we aim to measure a particular construct with the indicators,[3] for example treatment effectiveness or safety of care. The latter focus thus adds to the evaluation how many indicators we need to reliably measure that construct. Highly correlated indicators measure the same thing so retaining only one indicator could give enough information yet reduce the administrative burden for professionals. For instance, hospitals scoring well on collaboration between surgeon and plastic surgeon for invasive breast cancer in the study of Verheul et al[2] will likely also have good scores for ductal carcinoma in situ patients, as we are trying to measure collaboration and may question why this would be influenced by the type of patients. Similarly, we expect that the two indicators focused on irradical resection would be highly correlated. Previous studies have evaluated indicator sets in this way, lowering the administrative burden for data collection by retiring indicators.[4,5] Focusing on the set as a whole in these studies provided clarity on the construct measured by the indicators, such as psychosocial care[4] and treatment effectiveness,[5] which helps to improve their usefulness in daily practice but also in identifying possible missing quality of care dimensions such as patient-centeredness or equity. So, where evaluation of individual indicators within a set may address their measurement properties and feasibility, the evaluation of the indicator set as a whole shows whether the indicators collectively measure a specific construct without unnecessary duplication. Both provide distinct insights, but the focus of the evaluation that is chosen thus determines the criteria used.

What Are Relevant Criteria?

The criteria used in the evaluation need to be aligned with the type of indicators being evaluated. Indicators can be used for different purposes, well-established in the performance measurement literature ie, for accountability or for improvement.[6] Benchmarking is often used to show accountability, where measurement is focused on producing an accurate estimate of the performance, but will not illuminate how that performance is achieved or what is needed to improve it. An improvement indicator on the other hand is focused on providing a signal that can be connected to an actionable causal hypothesis (on what needs to change) and a subsequent test of change. An improvement indicator may thus collect just enough data to inform the best action forward as typically done in Plan-Do-Study-Act cycles,[7,8] rather than focusing on reliability and collecting data for multiple years as done for benchmarking. Given these differences, the criteria suitable for evaluating benchmarking or improvement indicators will likely also differ. Where benchmarking requires sufficiently reliable comparisons between organisations, local improvement may place greater emphasis on detecting meaningful changes over time and generating information to support appropriate improvement actions.

The indicators evaluated in the study of Verheul et al[2] are used for benchmarking hospitals. It thus makes sense that 3 of the 4 criteria are focused on ensuring the estimated difference in performance between hospitals is reliable and valid. Their evaluation thereby answers the question whether these indicators can support a valid comparison between hospitals, ie, can we reliably measure whether the quality of care in hospital A is different from that in other hospitals, but not other questions related to improvement. Such comparative hospital information can support various decisions as indicated by Verheul et al, such as helping patients choose a hospital. Patients may make more informed choices when they use the benchmarking indicators and receive care better aligned with their preferences. However, previous research has shown that public release of such comparative performance data had little impact on healthcare utilisation of patients who relied more on word-of-mouth and physician referral,[9,10] and of providers,[11] thereby pointing out that the information needs to be relevant for them to be used as intended. In this context, Verheul et al could have added usability as a criterion in their evaluation, which includes not only whether indicator information is presented in a clear and meaningful way but also if the information is relevant and useful to inform decisions of intended users, for instance by tracking the extent to which indicator information has been used by patients or other stakeholders to inform their decisions.

An evaluation of indicators focused on their use for improvement would likely include different criteria, for instance evaluating the extent to which there is still room for improvement by tracking the indicators over time, or their actionability eg, measuring the extent to which these indicators have resulted in desired actions and were effective in improving care. A very effective improvement intervention that resulted in all hospitals performing at a similar level, would then be evaluated quite differently when considering benchmarking indicators (rankability would be low) versus improvement indicators (very actionable).

Standardisation facilitates comparison across hospitals, whereas locally selected indicators may better reflect organisation-specific risks and improvement priorities. In other industries for instance, indicators are used as part of a company's risk management strategy where indicators are selected to monitor the most important organisation-specific risks and the extent to which these are controlled. Such locally selected indicators may be perceived as more useful but would also make an evaluation of indicator sets across companies very difficult (if not impossible). Collecting the same indicator data for all hospitals thus has advantages for the evaluation but may not capture the most important risks or processes for a specific hospital.

How and When Does the Evaluation Result in Further Actions?

An evaluation of the indicator set will only improve their usefulness if actions are taken based on the results of the evaluation. This means that we need to determine clear thresholds when such actions are needed. We may remove or change an indicator when it does not meet all criteria or use a different threshold eg, 80% of the criteria. Verheul and colleagues show that none of the indicators met all four

criteria and only 6 indicators met 3 of the 4 criteria,[2] but they are not explicit in their recommendations whether this means remaining indicators should be retired or changed. Previous studies evaluating indicator sets as a whole used a step-wise approach with clearly defined at each step when an indicator should be removed or changed, resulting in the final indicator set that was recommended for future use.[4,5] At the same time, these studies also identified what constructs were not measured by the indicator set and thereby identified areas where additional indicators were needed. This makes it clear how the evaluation informs selection of new indicators. In the study of Verheul et al, they report that indicators can be added during annual review by the scientific committee, based on expert opinion given new insights and clinical guideline revisions.[2] The link with experts and clinical guidelines is likely to facilitate the actual use of these indicators in daily practice and their relevance, but it remains unclear how the results of the evaluation by Verheul et al would be used in that process.

Particularly for benchmarking indicators, the extent to which indicators can reliably assess differences between hospitals may depend on methodological choices. This makes it an important consideration during the evaluation as it could result in refinement rather than removal of the indicator. Most funnel plots used in benchmarking hospitals implicitly use the mean or median performance to indicate acceptable performance, and Verheul and colleagues acknowledge that the average does not necessarily represent the most effective standard. We agree and would argue that benchmarking against a different standard eg, the desired goal such as achieving zero central line infections or detecting a (clinically) meaningful difference between hospitals might change the result of the evaluation. For instance, Bruinsma et al refined their indicator on total weight loss from 20% to 25% to improve the ability of detecting hospital differences.[5] Similarly, the use of a different method might also change the sensitivity of identifying deteriorating or significantly worse performing hospitals, thereby contributing to reliable assessment of differences between hospitals.[12,13]

Are All Stakeholders Involved?

An indicator set may be used by different stakeholders for different purposes and to answer different questions.[14] Verheul et al for example indicate that indicators may guide healthcare providers in improving care, assist patients in making informed choices about their care providers and support pay-for-performance strategies that incentivise high-quality care.[2] But what if these stakeholders do not agree about the relevance of a specific indicator, or about what constitutes a meaningful difference between hospitals? Dealing with conflicting stakeholder values is by no means straightforward, but an important consideration when designing an evaluation of an indicator set. Selection of indicators is often based on expert opinion and informed by clinical guidelines as done in the study by Verheul et al, even though it has been shown that patients value hospital reputation, physician experience and primary care recommendations as more important than guideline adherence or patient outcomes such as complications.[15] Comparative hospital information can be used by pay-for-performance schemes, but the effects on patient outcomes were shown to be small (at best).[16] This highlights how a single indicator set cannot be assumed to serve different stakeholders equally well. Therefore, the intended users and purposes should be specified before evaluation criteria are selected.

Conclusion and Final Reflections

The number of indicators has increasingly become an administrative burden, rather than a carefully chosen tool that is used to help achieve better and safer care. Systematic evaluation of indicators can help to improve their usefulness in daily practice, where we propose 4 key questions to ensure the evaluation approach is aligned with the intended use of the indicators and leads to meaningful decisions about changes in indicator sets. Having a set of indicators that is good for benchmarking does not necessarily make those good indicators for learning and improvement, so we need to strive for a good balance and target our improvement priorities carefully. In practice, we should aim for an efficient set of benchmarking indicators that provides reliable and valid information about the most relevant differences in quality of care between hospitals, complemented by locally selected indicators to monitor important risks and improvement indicators to signal what needs to change. Besides reporting the results of individual indicators, hospitals could also report on how they use these indicators to improve patient care in daily practice.

How can benchmarking indicators and their evaluation improve care? A possible theory could be: worse performance on the indicator than other hospitals –> internal and external pressures from consumers or inspectors –> incentive for healthcare professionals to investigate the causes –> develop interventions to prevent or reduce these causal factors –> improved care. Articulating such a theory makes clear how benchmarking will not automatically result in improvement, and why many researchers describe benchmarking as a starting point for improvement initiatives as it does not tell us what needs to be improved.[17] To illustrate, the indicator "Time from last surgery to start adjuvant therapy" in the study by Verheul et al[2] does not indicate the causes of the delay, and whether these causes occur more frequently in our hospital than in others. The assumption is also that delay equals worse quality of care, which in practice is not always the case as there may be a clear clinical indication for the delay and there could be a trade-off with other quality of care domains eg, less efficient but more effective care. This makes it difficult to draw conclusions on whether a change in this indicator equals improved care or not. The articulated theory does highlight how (external) pressures may induce the risk of organisations exhibiting strategic behaviour, resorting to 'managing the numbers' rather than true improvement.[18,19] Also, it highlights how the worse performance compared to others is needed as an incentive, thereby assuming that the large number of hospitals without such an incentive will not initiate any improvement actions. Whether and how performance measurement leads to improvement should be assessed empirically rather than

assumed to follow automatically from reporting performance data. Evaluation should not stop at asking whether indicators measure differences accurately; it should also consider the mechanism linking measurement to its intended use ie, how this information is interpreted, what decisions it informs, what actions follow and whether these actions contribute to better care.[20,21] This underscores the importance for the more balanced approach proposed above to ensure we make our indicators count in improving care for patients.

Disclosure of artificial intelligence (AI) use

Not applicable.

Ethical issues

Not applicable.

Conflicts of interest

Authors declare that they have no conflicts of interest.

Authors' contributions

Conceptualization: Perla J. Marang-van de Mheen, Jakko van Kampen, and Jop Groeneweg.
Writing–original draft: Perla J. Marang-van de Mheen.
Writing–review & editing: Jakko van Kampen and Jop Groeneweg.

Authors' affiliations

[1]Safety & Security Science Section, Delft University of Technology, Delft, The Netherlands. [2]Centre for Safety in Healthcare, Delft University of Technology, Delft, The Netherlands.

References

1. Zeegers M, Veenstra GL, Gerritsen G, et al. Perceived burden due to registrations for quality monitoring and improvement in hospitals: a mixed methods study. *Int J Health Policy Manag.* 2022;11:183–196. doi:10.34172/ijhpm.2020.96
2. Verheul EM, van der Linde M, Lingsma HF, et al. Comprehensive evaluation of quality indicators: analyzing the Dutch Breast Cancer Audit. *Int J Health Policy Manag.* 2025;14:8943. doi:10.34172/ijhpm.8943
3. Schang L, Blotenberg I, Boywitt D. What makes a good quality indicator set? A systematic review. *Int J Qual Health Care.* 2021;33:1-10. doi:10.1093/intqhc/mzab107
4. Xu D, Lewis T, Rurka M, Arling G. Approach to systematically examine the usefulness of quality measures in practice: Minnesota's nursing home quality indicators and scoring approach. *BMJ Qual Saf.* 2023;32:319-329. doi:10.1136/bmjqs-2021-014384
5. Bruinsma FFE, Nienhuijs SW, Liem RSL, Greve JWM, Marang-van de Mheen PJ. Refining the Quality Indicator Set from the Dutch Audit for Treatment of Obesity to Ensure Usefulness in Clinical Practice. *Obes Surg.* 2025;35:2616-2625. doi:10.1007/s11695-025-07898-2
6. Solberg LI, Mosser G, McDonald S. The three faces of performance measurement: improvement, accountability and research. *Jt Comm J Qual Improv.* 1997;23:135-147. doi:10.1016/s1070-3241(16)30305-4
7. Leis JA, Shojania KG. A primer on PDSA: executing plan-do-study-act cycles in practice, not just in name. *BMJ Qual Saf.* 2017;26:572-577. doi:10.1136/bmjqs-2016-006245
8. Etchells E, Ho M, Shojania KG. Value of small sample sizes in rapid-cycle quality improvement projects. *BMJ Qual Saf.* 2016;25:202-206. doi:10.1136/bmjqs-2015-005094
9. De Groot IB, Otten W, Smeets HJ, Marang-van de Mheen PJ. Is the impact of hospital performance data greater in patients who have compared hospitals? *BMC Health Serv Res.* 2011;11:214. doi:10.1186/1472-6963-11-214
10. Yahanda AT, Lafaro KJ, Spolverato G, Pawlik TM. A Systematic Review of the Factors that Patients Use to Choose their Surgeon. *World J Surg.* 2016;40:45-55. doi:10.1007/s00268-015-3246-7
11. Metcalfe D, Rios Diaz AJ, et al. Impact of public release of performance data on the behaviour of healthcare consumers and providers. *Cochrane Database Syst Rev.* 2018;9:CD004538. doi:10.1002/14651858.CD004538.pub3
12. Van Schie P, van Bodegom-Vos L, van Steenbergen L, Nelissen RGHH, Marang-van de Mheen PJ. Monitoring hospital performance with statistical process control after total hip and knee arthroplasty: a study to determine how much earlier worsening performance can be detected. *J Bone Joint Surg Am.* 2020;102:2087-2094. doi:10.2106/JBJS.20.00005
13. Kuhrij L, van Zwet E, van den Berg-Vos R, Nederkoorn P, Marang-van de Mheen PJ. Enhancing feedback on performance measures: the difference in outlier detection using a binary versus continuous outcome funnel plot and implications for quality improvement. *BMJ Qual Saf.* 2021;30:38-45. doi:10.1136/bmjqs-2019-009929
14. Marang-van de Mheen PJ, Vincent C. Measuring what matters: refining our approach to quality indicators. *BMJ Qual Saf.* 2023;32:305-308. doi:10.1136/bmjqs-2022-015221
15. Ellis RJ, Yuce TK, Hewitt DB, et al. National evaluation of patient preferences in selecting hospitals and healthcare providers. *Med Care.* 2020;58:867-873. doi:10.1097/MLR.0000000000001374
16. Mathes T, Pieper D, Morche J, et al. Pay for performance for hospitals. *Cochrane Database Syst Rev.* 2019; 2019(7):CD011156. doi:10.1002/14651858.CD011156.pub2
17. Van Schie P, van Steenbergen LN, Bodegom-Vos L, Nelissen RGHH, Marang-van de Mheen PJ. Between-hospital variation in revision rates after total hip and knee arthroplasty in the Netherlands: directing quality-improvement initiatives. *J Bone Joint Surg Am.* 2020;102:315-324. doi:10.2106/JBJS.19.00312
18. Franco-Santos M, Otley D. Reviewing and theorizing the unintended consequences of performance management systems. *Int J Management Rev.* 2018;20:696-730. doi:10.1111/ijmr.12183
19. Mannion R, Braithwaite J. Unintended consequences of performance measurement in healthcare: 20 salutary lessons from the English National Health Service. *Intern Med J.* 2012;42:569-574. doi:10.1111/j.1445-5994.2012.02766.x
20. Schang L, Morton A, DaSilva P, Bevan G. From data to decisions? Exploring how healthcare payers respond to the NHS Atlas of Variation in Healthcare in England. *Health Policy.* 2014;114:79-87. doi:10.1016/j.healthpol.2013.04.014
21. van Schie P, van Bodegom-Vos L, Zijdeman TM, Nelissen RGHH, Marang-van de Mheen PJ. Effectiveness of a multifaceted quality improvement intervention to improve patient outcomes after total hip and knee arthroplasty: a registry nested cluster randomised controlled trial. *BMJ Qual Saf.* 2023;32:34-46. doi:10.1136/bmjqs-2021-014472