

Look in the Mirror, Not Out the Window

In Favor of Internal Benchmarking

Defne Altan, AB,* Vanita Ahuja, MD, MPH,†‡ Cassandra M. Kelleher, MD,* David C. Chang, PhD, MPH, MBA*

Abstract: Over the past few decades, institutions have developed complex systems to compare themselves to others with the goal of improving healthcare quality. This process of comparison to others, called external benchmarking, has become the standard approach for quality improvement. However, external benchmarking is resource intensive, may not be flexible enough to focus on problems unique to individual institutions, and may lead to complacency for institutions ranking near the top of the quality bell curve for the measured metrics. Our singular focus on external benchmarking could also divert resources from other approaches. Here, we describe how the use of internal benchmarking, in which an institution focuses on improving their own processes over time, can offer unique advantages as well as offset the limitations of external benchmarking. We advocate for investment in both internal and external benchmarking as complimentary tools to improve healthcare quality.

Keywords: benchmarking, external benchmarking, healthcare quality, innovation, internal benchmarking, quality improvement.

A BRIEF INTRODUCTION TO BENCHMARKING

Over the past few decades, benchmarking has rapidly become a widely accepted means to pursue quality improvement in healthcare. As a concept, benchmarking originated in landscaping from “bench marks,” marks made on bedrock or “benches” to determine height based on reference to an established point.¹ Now used in the medical and business worlds, the most common approach to benchmarking entails objectively measuring and comparing outcomes across a group with the intent of improving performance. Because it compares individuals or organizations to each other, this approach is best described as *external* benchmarking. In contrast, internal benchmarking can be used to characterize outcomes for a single performer to determine best practices. Internal benchmarking can mean comparing across internal departments as well as tracking changes of an individual organization over time. External benchmarking is widely used throughout medicine today. The purpose of this commentary is to advocate for increased adoption of internal benchmarking, which is a valuable, often-underutilized tool that can be complimentary to existing quality improvement systems.

From the *Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA; †Department of Surgery, Veterans Affairs Connecticut Healthcare System, West Haven, CT; and ‡Department of Surgery, Yale School of Medicine, New Haven, CT.

Disclosure: The views expressed in this article are those of the author and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

Reprints: David C. Chang, 165 Cambridge Street, Suite 403, Boston, MA 02114. E-mail: dchang8@mgh.harvard.edu.

Copyright © 2022 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Annals of Surgery (2022) 3:e184

Received: 23 February 2022; Accepted: 14 June 2022

Published online 19 July 2022

DOI: 10.1097/AS9.000000000000184

EXTERNAL BENCHMARKING

External benchmarking in surgery first began in 1991 in response to higher morbidity and mortality in the Veterans Affairs Health System²; this was later expanded into the American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP), an open-subscription program for nonfederal hospitals. Participation in ACS NSQIP has been associated with reduced adverse events after surgery.³ Due to inherent deficiencies in reporting specific procedural-related complications, many surgical subspecialties then started their own external benchmarking programs—such as the Society of Thoracic Surgery, Society for Vascular Surgery, and the American Society for Metabolic and Bariatric Surgery—to highlight clinical outcomes and metrics relevant to their specialties.⁴ The advantages of external benchmarking are manifold. For example, it is uniquely equipped to address situations where no known norm or generally accepted standard exists. It can help provide context for methods and strategies to improve, directing interventions towards areas of weakness where they might have maximum benefit. A large amount of data from external benchmarking also make trends easier to track and understand.

However, external benchmarking also has limits. They can be seen at each phase of its operation: during setup, throughout measurement, and after measurement is complete.

Data Burden and Resource Constraints

In the setup phase, external benchmarking usually requires the collection of large quantity of data up front.^{5,6} To meaningfully compare across individuals in a group, contextual data about each individual are necessary. For example, if the goal is to characterize body habitus, it is not valid to simply compare weights without knowing corresponding heights. Weight, then, needs to be “risk adjusted” for height. For more nuanced comparisons, adjustment may be needed for age, sex, and several other factors. This is why we need age- and sex-specific growth charts. When larger groups are compared, more nuances must be accounted for. On the nationwide scale of patient outcomes, these comparisons often require massive time and resource investment. For example, NSQIP collects over 60 preoperative variables to adjust for operative risk.⁵ To meet the data burden of interhospital comparison via NSQIP, participating hospitals must delegate personnel, set aside annual funds for technical support and data

analysis, and conduct follow-up on all participating patients as well as routine inter-reliability audits.⁶ However, hospital resources are finite, and time, energy, and effort reserved for external benchmarking may paradoxically shift the focus away from quality improvement. In addition, adequate telecommunication infrastructure for data-sharing must be in place to accommodate external benchmarking, which can be challenging for resource-limited settings.

Metric Choice

In the measurement phase, external benchmarking must rely on metrics shared by all its members. However, these common denominators may not be applicable or even relevant to every member of the group. For example, pressure ulcers are an often-used patient safety metric, and higher rates of pressure ulcers can indicate poorer inpatient care. Yet, for high-volume specialty hospitals with patients who tend to stay only a few days, pressure ulcer rate becomes less helpful as a quality metric. *What is measured, is managed*; a focus on broadly applicable metrics might lead a hospital to misdirect resources toward improving other hospitals' problems rather than their own. On the opposite side of the spectrum, rare but important events, the so-called Never Events, are difficult to describe and understand through external benchmarking. When rates hover around zero, it is impossible to meaningfully benchmark. One example is wrong-sided surgery, which is catastrophic but extremely rare. An absence of metrics surrounding rare but important outcomes can divert resources away from solving these problems.

Innovation Gap

After measurement is complete, data from external benchmarking can sometimes be counterproductive to the ultimate goal of quality improvement. Those who score near the top of the curve may believe they have little reason to improve on that metric, since they are performing better than their peers. This can encourage complacency and offer less incentive for innovation in areas where performance is already relatively "good." In the few years before 2014, General Motors was outpacing its competitors in global sales, even while an ignition switch defect in one of its models was slowly but surely causing fatal car crashes.⁷ Here, above-average performance made it easier to overlook a crucial safety issue. In this context, conflicting data from different external benchmarking systems can lead to further confusion and inaction. External benchmarking can also sometimes facilitate groupthink. The practice of learning from other performers (often the high performers) may yield useful and important insights. However, this can encourage all members of the group to take similar approaches to solving a problem that are not necessarily tailored to their given situation; what works for some may not work for all. Additionally, data obtained from external benchmarking may turn the focus towards marketing rather than quality improvement,⁸ especially for those at the top of the curve.

INTERNAL BENCHMARKING

Despite the popularity of external benchmarking, internal benchmarking, first pioneered by Dr Ernest Codman at MGH, has a longer history in medicine. Internal benchmarking drives quality improvement grounded in the context of each particular hospital, bolstered by many strengths. First, internal benchmarking requires significantly less data. Because each hospital serves as its own control, complex "risk adjustment" can be avoided. There is less variation across time for one person, than across people in a group. This reduced data burden can help overcome the barriers many institutions face in implementing quality improvement systems. Second, internal benchmarking also allows one to select

unique metrics suited to their individual setting. A hospital with a recent wrong-site surgery can focus on benchmarking relevant metrics (such as the use of time-outs and checklists) to address its unique system failings, while other hospitals can continue to focus on different patient safety metrics. Third, internal benchmarking provides everyone with motivation to improve regardless of where they perform relative to others. Above-average performers will continue to have incentive for quality improvement and innovation; since there is no possibility of being "too good," one can always be better than yesterday. In fact, to the extent that we want to foster cultures of continuous self-improvement, a system of internal benchmarking will be more likely to achieve that goal than external benchmarking. Internal benchmarking encourages one to look inward for solutions. Finally, it will also avoid the risk of high performers advertising at the expense of their competitors; instead, internal benchmarking will encourage advertising that is solely based on *that hospital's* performance from 1 year to the next. And a hospital that is constantly improving might be more attractive to patients than a hospital that merely rests on its laurels.

Despite its strengths, internal benchmarking also has limitations. For example, internal benchmarking alone may not provide the context needed to understand how serious a problem is and prioritize interventions accordingly. With only internal data, a hospital may direct its resources towards perceived problems, not toward domains where it is performing more poorly than other hospitals. Second, a focus on internal processes and outcomes through internal benchmarking may prevent hospitals from implementing outside solutions. Third, internal benchmarking often comes with a smaller sample size within just one department or hospital, which makes it more difficult to detect trends.

Internal benchmarking is successful in many areas of quality improvement in medicine. For example, it has been used to track nursing productivity over time,⁹ to measure internal variation before and after healthcare interventions and within a yearly cycle,¹⁰ and to track metrics like patient satisfaction and patient-reported outcomes.¹¹ Internal benchmarking is particularly helpful when it is difficult to adjust for unique patient population characteristics, such as malnutrition or asbestos exposure. In some cases, internal and external benchmarking can be used in tandem to fully understand clinical outcomes, both in context of the organization as well as the larger healthcare landscape.^{10,11}

We advocate for a mixed model approach to benchmarking, where internal and external benchmarking are used in a complementary way rather than the exclusion of one or the other. In our hurry to look out the window to see what others are doing and learn from them, we should also not forget to first look in the mirror.¹²

REFERENCES

1. Zairi M, Leonard P. *Practical Benchmarking: The Complete Guide*. Dordrecht: Springer Netherlands; 1996, pp. 22–27
2. Ingraham AM, Richards KE, Hall BL, et al. Quality improvement in surgery: the American College of Surgeons National Surgical Quality Improvement Program approach. *Adv Surg*. 2010;44:251–267.
3. Cohen ME, Liu Y, Ko CY, et al. Improved surgical outcomes for ACS NSQIP hospitals over time: evaluation of hospital cohorts with up to 8 years of participation. *Ann Surg*. 2016;263:267–273.
4. Epelboym I, Gawlas I, Lee JA, et al. Limitations of ACS-NSQIP in reporting complications for patients undergoing pancreatotomy: underscoring the need for a pancreas-specific module. *World J Surg*. 2014;38:1461–1467.
5. Anderson JE, Lassiter R, Bickler SW, et al. Brief tool to measure risk-adjusted surgical outcomes in resource-limited hospitals. *Arch Surg*. 2012;147:798–803.
6. Hospital Requirements. American College of Surgeons: ACS NSQIP. 2021. Available at: <https://www.facs.org/quality-programs/acs-nsqip/joinnow/hospitalreq>. Accessed October 10, 2021.

7. Basu T. Timeline: a history of GM's ignition switch defect. Npr.org. 2021. Available at: <https://www.npr.org/2014/03/31/297158876/timeline-a-history-of-gms-ignition-switch-defect>. Accessed October 10, 2021.
8. Neuman HB, Michelassi F, Turner JW, et al. Surrounded by quality metrics: what do surgeons think of ACS-NSQIP? *Surgery*. 2009;145:27–33.
9. Morin. Challenging the status quo to innovate the future of nurse productivity and benchmarking. *J Obstetric Gynecol Neonatal Nurs*. 2020;49:S82–S82.
10. Al-Kuwaiti A, Homa K, Maruthamuthu T. A new performance improvement model: adding benchmarking to the analysis of performance indicator data. *Joint Commission J Qual Patient Saf*. 2016;42:462–465.
11. Warnakulasuriya SR, Patel RC, Singleton GF, et al. Patient-reported outcomes for ambulatory surgery. *Curr Opin Anaesthesiol*. 2020;33:768–773.
12. Puckett J, Siegel P. Looking in the mirror. *J Business Strategy*. 1997;18:12–16.