Lessons for philanthropy

# What philanthropy can learn from healthcare performance benchmarking

By Ben Harder

GuideStar by Candid.

# Introduction

Performance measurement is an important dilemma facing the philanthropic sector. Charities with similar missions may operate in parallel but lack shared, validated yardsticks with which to assess their social impact and efficiency. Identifying a set of shared performance measures will enable comparisons among organizations and their respective initiatives, first, by helping charities identify opportunities to improve, and, second, by helping both charities and donors maximize their impact through efficient allocation of philanthropic resources.

Charities need not develop a novel approach to solving this challenge. Other sectors facing similar needs have made notable strides in performance measurement. Cross-disciplinary insights from these efforts can inform and accelerate philanthropy's development of measures of impact and efficiency.

Healthcare, in particular, offers a useful analogy. Considerable recent progress in healthcare performance measurement invites emulation. Performance measurement in healthcare is instructive to philanthropy for several reasons. First, many organizations that provide healthcare services, including the majority of U.S. hospitals and many health insurers, are nonprofits. Consequently, healthcare leaders often have direct personal experience with the challenges facing both sectors.

Second, like philanthropies, healthcare providers have a social mission, in that they aim to improve the lives of individuals and the vitality of communities.

Third, both healthcare organizations and charitable organizations engage in activities that accrue societal benefits beyond those they directly provide to individual recipients of services.

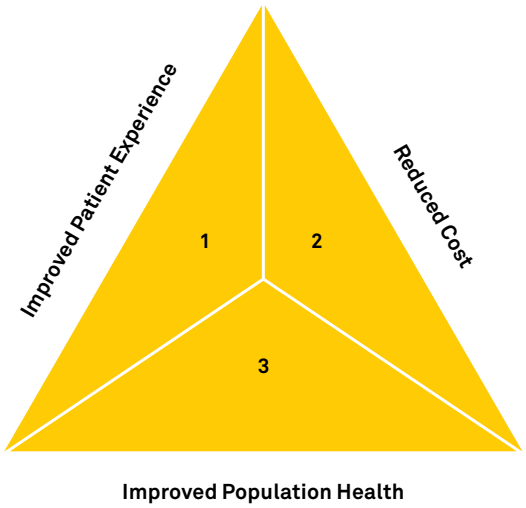
It is helpful to understand the healthcare sector's measurement efforts in the context of recent health policy reforms. As codified in laws such as the 2010 Patient Protection and Affordable Care Act (aka Obamacare) and the 2015 Medicare Access and CHIP Reauthorization Act, recent health policy reforms have increasingly encouraged healthcare providers such as hospitals and doctors to focus on three intertwined aims:

1. to improve the quality of care for the sick;

2. to reduce the societal cost of healthcare; and

3. to increase the health of populations so that fewer people will become sick and fewer costly services will be required.

This so-called "Triple Aim," a term popularized by Dr. Don Berwick, former administrator of the Centers for Medicare and Medicaid Services (CMS) and co-founder of the Institute for Healthcare Improvement, is widely viewed as an objective that all healthcare stakeholders should seek. To achieve progress toward the Triple Aim, CMS and other healthcare stakeholders have increasingly used quantitative measurement—of quality, cost, and value—paired with public transparency. For example, CMS, which pays for a significant fraction of U.S. healthcare services, instituted a "Value-Based Purchasing Program" to track various measures of hospitals' performance and financially reward hospitals that meet certain benchmarks, such as reducing hospital-acquired infections and improving patient satisfaction.




Philanthropy, too, is undergoing reform, albeit of a less formal sort. The rise of impact-oriented philanthropy (as exemplified in the recent writings of William MacAskill and Peter Singer) as well as the popularity of impact investing reflect an increasing focus among some philanthropists on quantifying and maximizing their "returns" on charitable "investments." This trend puts pressure on donor-dependent nonprofits to incorporate quantitative evidence of impact and efficiency into their fundraising narratives.

# Measuring what matters

Measuring philanthropic return, however, is challenging. As the social scientist W. B. Cameron famously wrote, "Not everything that can be counted counts, and not everything that counts can be counted." (The quote is often misattributed to Einstein.)

Despite Cameron's truism, stakeholders tend to be drawn to performance benchmarks based on activities that are readily quantifiable, such as the cost to raise a dollar or the share of a charity's spending that goes toward administrative expenses (overhead).

An unintended consequence is that hard-to-measure domains of performance may go underappreciated and undervalued by stakeholders. For example, the relative ease with which donors can compare charities' proportion of spending on administrative overhead has led to an overemphasis on this ratio, while obscuring potentially meaningful—but unmeasured—differences in results.

It's possible to devise results-oriented performance measures that could be used for organization-level comparisons. Doing so in philanthropy will require (1) shared definitions of measures that matter and (2) data collection, validation, and reporting to a wide range of stakeholders. There is an array of examples from healthcare where stakeholders have collaborated in order to meet these two conditions, with the result that performance benchmarking has become increasingly robust and unintended consequences have been mitigated or avoided.

# Performance measurement in healthcare

Healthcare performance measurement is not the domain of any single organization. Rather, multiple measurement-focused organizations coexist in an organic and fluid ecosystem. While the philanthropic sector might take an alternative—and potentially tidier—approach, it's worth understanding the current landscape of measurement in healthcare so that useful concepts can be borrowed and critical functions can be identified.

A variety of healthcare organizations play critical roles. These include the National Quality Forum (NQF), a semi-public nonprofit that acts as a clearinghouse of validated healthcare performance measures; various clinical data registries, each typically administered by a national nonprofit organization that represents practitioners of a particular medical specialty, such as the Society of Thoracic Surgeons (STS), which comprises heart surgeons and lung surgeons; consumer-service organizations such as *Consumer Reports*, Consumers' Checkbook, Leapfrog Group, and my employer, *U.S. News & World Report*; the federal agency Centers for Medicare and Medicaid Services (CMS); and other organizations.

## Defining measures that matter

The NQF employs a multi-stakeholder, consensus-based process to evaluate and endorse measures that its member organizations have developed for use in healthcare performance benchmarking. Its members include universities, academic medical centers, various consortia, and federal agencies such as the Agency for Healthcare Research and Quality.

In all, the NQF has endorsed more than 400 quality measures. While NQF ratification of a measure carries no regulatory implications, endorsed measures may be used for regulatory purposes by CMS. In addition, hospitals, insurers, and third parties often use NQF-endorsed measures to track, improve, and reward healthcare providers' performance.

## Collaborative benchmarking

Clinician-led collaborative efforts have been largely positively received by healthcare stakeholders. Heart surgeons, transplant surgeons, and pediatric doctors who treat cystic fibrosis are among the healthcare

professionals who, within their respective peer groups, have devised consensus-based performance measures and created systems for data collection and collaborative benchmarking. Such efforts are typically referred to as "clinical registries," because each case treated by any participating clinician is registered in a common database.

Clinical registries collect detailed data on an ongoing basis from clinical-care providers such as hospitals, and use these data to measure participants' performance. Typically, doctors and hospitals that participate in these registries receive periodic reports that compare their individual performance to peer benchmarks across a range of measures.

For example, a hospital that contributes data to the Adult Cardiac Surgery Database, a registry maintained by STS, receives a biannual performance report comparing its own patients' death rates and complication rates to those of hospitals that treated patients with similar severity of heart disease. Such reports enable participating hospitals to identify areas for improvement and to select high-performing peers on which to model their self-improvement efforts.

Providers have achieved these performance gains in spite of the fact that most registries are not transparent to external stakeholders such as patients; **hospital- and physician-level performance reports generally are not made publicly available.** Instead, the benchmarking results are held confidentially within the collaborative, in order to minimize the risk of unintended consequences that might result from public misinterpretation of the data.

**Consequently, patients (and donors) cannot use these benchmarking services for informed decision making,** so society is unable to capture the full potential benefit of registries' measurement efforts.

## Public benchmarking

Some registries, including STS (sts.org) and the Scientific Registry of Transplant Recipients (srtr.org), have begun to issue public reports that are available to prospective patients and other external stakeholders. More registries are considering doing so. In most cases, public disclosure is optional—a voluntary decision made by each participating hospital or doctor.

Meanwhile, another group of stakeholders has tackled performance measurement with the express goal of providing benchmarking services to the public. One such effort is led by the Leapfrog Group, a nonprofit that surveys hospitals to obtain data pertinent to patient safety, such as whether a hospital adheres to certain established best practices. Leapfrog publishes a free online tool, Hospital Safety Grade (hospitalsafetygrade.org), that benchmarks hospitals using NQF-endorsed measures and provides an A-to-F summary grade for each hospital.

Another consumer-facing benchmarking initiative is the Best Children's Hospitals project, which I oversee as an employee of *U.S. News & World Report*, the project's sponsor. *U.S. News* annually surveys more than 100 children's hospitals using a detailed data-collection instrument that was collaboratively designed by *U.S. News*, its data contractor RTI International, and dozens of volunteer pediatric physicians who make up several *U.S. News*/RTI-convened working groups. The resulting data set is analyzed to produce an array of performance benchmarks and indices (i.e., rankings) that *U.S. News* publishes and makes freely available to the public.

Patients, their families, referring pediatricians, researchers, hospital leaders, and donors use these indices as tools to help inform their decisions. More than five million users per month access *U.S. News*'s online provider profiles, including the indices of Best Children's Hospitals. In part because of this broad reach, *U.S. News*'s public benchmarking has considerable salience among diverse stakeholders.

## Government-led benchmarking

CMS, which is a federal agency, has assumed a high-profile role in healthcare performance measurement; it publishes ratings of hospitals, nursing homes, health insurance plans, home health agencies, and other healthcare service providers. However, those efforts have been criticized by stakeholders and questioned by investigative journalists.

Federal initiatives in educational benchmarking, such as the College Scorecard, have been similarly controversial. While it's conceivable that policymakers might come to view philanthropic measurement as an appropriate function of a federal agency such as the Internal Revenue Service, the odds appear low. Moreover, the government's track record suggests that agencies may not provide optimal platforms for performance measurement.

# The outlook for benchmarking in philanthropy

Who or what will measure and publicly benchmark the performance of philanthropic organizations?

To date, initiatives to benchmark nonprofits' performance remain nascent. One of the most ambitious efforts, led by Charity Navigator, has produced measures of "financial health" and "accountability and transparency" for an array of nonprofits. However, Charity Navigator's methodology for benchmarking a third, vital dimension of performance, "results reporting," remains in development. In the absence of a mechanism to gauge philanthropies' comparative results, some stakeholders have concerns that measures of financial efficiency will take on outsized importance in the eyes of donors and nonprofit leaders, perpetuating the "Overhead Myth."

Arguably, what philanthropy needs in order to benchmark results is a registry like those that are flourishing in healthcare. With support from a registry, nonprofits could (a) collaboratively define suitable performance measures, (b) collect internal measurements, (c) pool those data alongside data from peer organizations, and (d) receive the benchmarking services they need to track, improve, and communicate their organizations' results.

An effective platform for such a philanthropic registry would be a neutral, trusted third party that has experience compiling and managing organization-level nonprofit data. Its charge would be to convene peer groups of nonprofits, receive data voluntarily submitted by those organizations, and perform benchmarking services for those organizations. It may or may not be advisable for a philanthropic registry to publicly reveal its benchmarking findings for individual participating nonprofits. While public transparency is generally desirable, it could inhibit some organizations from participating in the registry, which might limit the registry's societal value. Therefore, at least initially, public disclosure might be left to the discretion of each participating nonprofit, as it typically is in the context of healthcare registries. Universal public reporting to external stakeholders might warrant subsequent consideration, once participating nonprofits grow sufficiently comfortable with nonpublic benchmarking.

Multiple philanthropic organizations possess useful capabilities and could play important roles in this push. For example, data platforms such as Candid, formed when Foundation Center and GuideStar joined forces

in February 2019, could provide a technological backbone to support data collection and could contribute expertise in establishing data standards. Ratings organizations such as Charity Navigator might perform data analysis and issue benchmarking reports to participating nonprofits and/or to the public. Media outlets such as ProPublica might be enlisted to interpret broader patterns in the data and highlight potential policy implications.

Here, too, healthcare offers an encouraging vision of what the philanthropic sector could achieve. Many independent healthcare registries now flourish alongside one another, each having developed its own area of medical expertise and its own set of stakeholder-approved performance measures. By the same token, religious nonprofits, for example, may choose to use different measures than educational institutions, and each peer group of nonprofits may choose to create its own philanthropic registry. Moreover, competition in healthcare among benchmarking organizations ensures that no single organization's definition of "quality" becomes dogmatic. Similarly, there may be room for pluralism in benchmarking philanthropy.

That said, certain functions, such as establishing data standards and aggregating data, will be most efficiently performed if they are centralized. In healthcare, federal agencies play a central role in this respect. In philanthropy, it seems that a nonprofit, or multiple collaborating nonprofits, will have to fill this gap themselves.

# Conclusion

Applying key lessons from recent advances in healthcare performance measurement, philanthropic stakeholders may benefit from one or more organizations that individually or collectively serve these public functions:

1. Standard-setting. Facilitate a consensus-based process to define an inventory of meaningful performance metrics that apply, in varying combinations, to as wide a range of charitable organizations as possible.

2. Data collection. Collect data from participating charities that enable the calculation of agreed-upon performance metrics. Relevant data may be collected both directly, such as by voluntary survey of charities, and indirectly, such as through IRS Forms 990.

3. Comparative analysis. Calculate key performance metrics for each charity to which a metric is relevant, and compare similar organizations to derive performance benchmarks.

4. Reporting to stakeholders. Report benchmarks and organization-level performance (a) in a confidential manner that enables charitable organizations to undertake performance-improvement efforts and, (b) if and when appropriate, report in a publicly transparent manner to enable external stakeholders to compare the performance of organizations with similar missions.

Providing these services to the philanthropic ecosystem could drive improvements in impact, efficiency, and donor engagement. Those three objectives are a worthy Triple Aim.

**About the author**

Ben Harder is managing editor and chief of health analysis at *U.S. News & World Report*, where he oversees a portfolio of decision-support tools for healthcare consumers. These free tools, available at usnews.com/health and used by millions of consumers per month, include hospital rankings and ratings, nursing home ratings, and a searchable directory of nearly every physician practicing in the United States.

Harder's writing has appeared in the *Washington Post, USA Today, Los Angeles Times, Chicago Tribune, Science News, Journal of the American Medical Association, and British Medical Journa*l as well as the anthology *Best American Science Writing 2005*.


https://doi.org/dkcq