

ARTICLE

Measuring and Improving Quality in the US: Where Are We Today?

Elizabeth A. McGlynn, PhD

In the 50 years since the American Board of Family Medicine (ABFM) was established, the United States has gone from a shared perspective that high-quality care was being routinely delivered to becoming aware of the significant and pervasive problems with quality. Efforts to stimulate improved quality have included public reporting, pay for performance, and value-based purchasing. In addition, maintenance of certification, systematic reviews of research, practice guidelines, electronic health records, and quality improvement programs have offered support for different dimensions of quality. Despite these programs and infrastructure, there is little evidence that quality has improved systematically in the United States. There are areas in which quality is better but many other areas in which quality has remained the same or even worsened. The focus on financial incentives as a primary tool for motivating improvement may not be productive and there is little evidence from research that quality varies with payment or incentives. Quality is a systems issue and requires system solutions. The ABFM has had a long commitment to assessing quality and has an opportunity to lead the way in reimagining quality measurement and assessment. (J Am Board Fam Med 2020;33:S28–S35.)

Keywords: Certification, Pay for Performance, Quality Improvement, Quality of Health Care, Value-Based Purchasing

The American Board of Family Medicine's (ABFM's) 50th anniversary offers an opportunity to reflect on the United States' journey in measuring and improving quality. Fifty years is a long time and the blink of an eye for medicine and for the quality journey. Fifty years ago, Denton Cooley implanted the first artificial heart and AIDS is believed to have first migrated to the United States. Advances in science and technology in that year included the Apollo 11 moon landing and walk, the first Concorde test flight, and the first automated teller machine. Medicare and Medicaid were only a few years old when Richard Nixon declared, "We face a massive crisis in health care costs."¹

A Brief, Selected History of Quality Measurement and Performance

Quality of care has been the subject of research and policy debates for a long time as illustrated by selected examples. Florence Nightingale, a leading statistician of her time, documented that unsanitary conditions in military hospitals were the major reason for preventable deaths among soldiers fighting in the Crimean War in 1853. In 1917, Ernest Codman, a pioneer in measuring and classifying medical errors and in assessing the long-term outcomes of surgery, publicly reported on surgical outcomes at Massachusetts General Hospital and called on all hospitals to do the same. A study of care received by members of the Teamsters Union in New York found that 57% of hospital care met recommended standards.² Lyons and Payne,³ using a novel, peer-reviewed episode of illness method, found that postdischarge care in Hawaii met standards 41% of the time. In 1976, Rhee⁴ reported on variations in quality among 454 physicians in 18 specialties ranging from 45% for stroke to 91% for cesarean section.

Research on quality has examined both overuse (delivering care for which the expected benefits do not outweigh the expected risks) and underuse (failure to deliver care that has been shown to be

This article was externally peer reviewed.

Submitted 28 October 2019; revised 5 May 2020; accepted 5 May 2020.

From the Kaiser Permanente Research, Kaiser Permanente, Pasadena, CA; Kaiser Permanente Bernard J. Tyson School of Medicine, Pasadena, CA.

Funding: None.

Conflicts of interest: None.

Corresponding author: Elizabeth A. McGlynn, PhD, Kaiser Permanente School of Medicine, 100 S. Los Robles, 3rd Floor Pasadena, CA 91101 (E-mail: Elizabeth.A.McGlynn@kp.org).

beneficial). Wennberg⁵ reported on variations in the rates with which common surgical procedures were performed across the United States in the Medicare population. Brook and colleagues⁶ at RAND and University of California, Los Angeles (UCLA) led studies of the appropriateness with which diagnostic and surgical procedures were used. The early studies using this method⁷⁻¹³ found that about one third of procedures were not clinically appropriate, that is, the potential health risks to patients undergoing the procedures were equal to or greater than the potential health benefits.¹⁴ However, there was no relationship between appropriateness and the rates of utilization.¹⁵⁻¹⁷ Similar rates of inappropriate procedure use were found in the United Kingdom,¹⁸ Israel,¹⁹ Canada,²⁰ and Sweden²¹ with very different models of health care financing, and overuse and underuse could co-occur within the same county.²²

In the late 1990s, a literature review found consistent and pervasive deficits in quality.²³ Despite this body of research, no strong signal had yet emerged that the United States had suboptimal quality. This began to change in the late 1990s and early 2000s with a series of reports and publications. In 1999, the Institute of Medicine (IOM), renamed the National Academy of Medicine in 2015, published *To Err is Human*, raising attention about the problem of errors in the delivery of medical care.²⁴ The committee estimated that up to 98,000 people died in the United States each year as a result of medical errors. The IOM followed this with *Crossing the Quality Chasm*, which provided a broader framework for assessing and addressing quality problems.²⁵

Along with the IOM reports, 4 articles from a national study also contributed to breaking through the perception barrier. The study enrolled a random sample of 6700 adults in 12 geographic areas broadly representative of the United States.²⁶ Participants completed a health history survey and gave investigators permission to obtain medical records from all clinicians and institutions from which they had received medical care in the 2 years before enrollment. Using methods previously developed at RAND, the team developed 439 quality measures for 30 acute and chronic conditions representing the leading causes of death and illness, and preventive care. Scores were constructed by identifying all quality indicators for which an individual was eligible and counting the number of those indicators that were

received or offered. The first article, published in 2003, reported that American adults were receiving about 55% of the recommended care for the leading causes of illness and death.²⁶ Performance on the quality indicators was similar for preventive, acute, and chronic care but varied by condition. The second article reported that overall quality ranged from 51% in Little Rock, Arkansas to 59% in Seattle, Washington.²⁷ Quality varied by condition and no community was always the best or worst on any of the dimensions examined. The communities varied in population growth rates, average income, poverty levels, rates of uninsured, hospital beds and physicians per 1000 population, and penetration of managed care, but there was no discernible relationship between these economic factors and quality. The third article reported that differences in quality between demographic groups were smaller than the gap between observed performance and optimal performance.²⁸ The fourth article reported that children in the households participating in the national study of adults were receiving 47% of recommended ambulatory care.²⁹ The best performance was observed for acute problems (68%) followed by chronic problems (53%) and then preventive care (41%). Quality for specific conditions ranged from 92% for upper respiratory infection to 34% for adolescent preventive care.

Policy Responses to Address Gaps in Quality

Different approaches to addressing gaps in quality have been undertaken since then. Notably, all these approaches rely on quality measurement as the basis for reporting, incentives, or contracting. These policy approaches were built on top of existing mechanisms for ensuring quality such as professionalism, licensure, and board certification.

Public Reporting of Quality Performance

Public reporting may affect quality through transparency, consumer choice, and reputation. Making quality performance results publicly available provides transparent information about variations in performance. Transparency enables consumers or their agents to use quality reports to choose health plans, hospitals, and doctors. Transparency and consumer choice also increase the motivation of clinicians, hospitals, and systems to improve performance because of concerns about reputation.

In 1986, the Health Care Financing Administration (now the Centers for Medicare and Medicaid Services [CMS]) produced national public reports on hospital mortality for 17 medical and surgical conditions.³⁰ New York state followed with reports on hospital mortality rates for coronary artery bypass graft surgery.³⁰ Since then public reports on quality have been released at various levels in the system—hospital, health plan, medical group, nursing home, physician. The Agency for Healthcare Research and Quality (AHRQ) commissioned a comprehensive review of the literature on public reporting published between 1980 and 2011. Public reporting was associated with an increase in quality improvement activities, some improvements in some quality measures, but little evidence that it affected choice of health care providers by patients.³¹ The authors found considerable heterogeneity in outcomes and moderate quality evidence making it difficult to draw definitive conclusions. A more recent review examining the impact of public reporting on clinical outcomes found mostly a positive effect on mortality (risk ratio, 0.85; 95% CI, 0.79 to 0.92) although the authors noted considerable heterogeneity among studies.³² Consumer choice of health plans, hospitals, and physicians is a complex task which may explain why transparency has not been associated with major changes in consumer choices.³³

Pay for Performance

Fee-for-service payment is generally believed to incentivize clinicians and systems to provide more services than necessary (overuse). Capitation payments raise concerns that clinicians and systems are incentivized to withhold needed care (underuse). Pay for performance (P4P) was developed to reward quality within either type of health care payment system. In P4P, payments to clinicians, hospitals, and systems can be adjusted by adding a quality-based incentive payment (or making a portion of overall payment contingent on quality performance). Private purchasers were early adopters of this approach. For example, in 2001 the California Integrated Healthcare Association created one of the first P4P programs for physician groups. Blue Cross Blue Shield of Massachusetts introduced the Alternative Quality Contract in 2009 with a similar focus on physician groups. CMS collaborated with Premier on the Hospital Quality Incentive Demonstration Program from 2003 to 2009 targeting hospital care for 3 conditions: acute myocardial infarction, heart failure,

and pneumonia. CMS also ran the Physician Group Practice Demonstration. States have tried various P4P programs for Medicaid and Children's Health Insurance Program providers. The design of P4P programs varies and can include both positive and negative incentives (bonuses and penalties).

Research on P4P has found mixed results. A study of the Premier hospital P4P program found early short-term effects that converged with the control group about 5 years into the program.³⁴ A review of P4P similarly found short-term (2 to 3 years) positive effects on processes of care with longer term effects uncertain.³⁵ Positive studies tended to be those in the United Kingdom or in areas with very low baseline performance. No consistent effects were found for intermediate or long-term health outcomes. Although considerable heterogeneity exists among programs, the review concluded that this does not change the mixed assessment of success.³⁵ Although the literature on P4P finds mixed or modest effects, 1 US³⁶ and 1 UK³⁷ study found that performance declined after selected P4P incentives were removed.

Value-Based Purchasing

This extends P4P with a more explicit focus on simultaneously assessing quality and spending. Bundled payments—a lump sum payment for a specific episode of care for a condition—were an early approach to value-based purchasing; private purchasers were early adopters. The approach has been extended to populations, such as those engaged in Medicare's Accountable Care Organizations. Another variation on this approach is value-based insurance design, in which patient cost sharing is adjusted to incentivize high-value care (or disincentivize low-value care).

Research on these programs has found similar marginal effects.³⁸ An evaluation of Medicare's Hospital Value Based Purchasing Program found no differences in process of care, patient experience, or mortality.³⁹ The Medicare Payment Advisory Commission has recommended combining the 4 hospital quality payment programs into a single Value Incentive Program to bring the programs into alignment with the Commission's principles which include a broad measures of overall performance rather than focus on a limited number of conditions.⁴⁰

Researchers and reviewers have noted a variety of issues that might explain the modest effects

associated with quality payment programs such as: design and choice of quality measures, magnitude and design of incentives, method of attributing patients to providers, and the effect on vulnerable populations.

Enablers for Quality: Additional Efforts to Support Quality

The previous section highlighted specific programs that use quality measures and systematic assessments of performance to incentivize quality improvement. A parallel effort is developing needed infrastructure or support for improved quality.

Maintenance of Certification

Time limited board certification reflects a belief that knowledge and skills should be periodically reevaluated because the science of medicine continues to evolve. The ABFM program includes 3 components relevant to quality: self-assessments of knowledge, a secure examination, and performance improvement modules. This approach combines mechanisms to raise awareness of potential gaps in knowledge, a high-stakes examination to demonstrate knowledge, and practice-specific exercises to examine and improve performance.⁴¹

Systematic Evidence Reviews

Systematic reviews are one approach to dealing with the challenge of the volume, variable study designs, and quality of published research. Such reviews summarize what is known about the potential benefits and harms of drugs, devices, and other health care services. An IOM committee developed standards for systematic reviews to encourage high quality reviews.⁴² AHRQ supports the Evidence-based Practice Center program, a group of 12 centers that conduct systematic reviews for both the federal government and other entities.⁴³ The Cochrane Collaboration is another source of high-quality systematic reviews.⁴⁴

Practice Guidelines

Practice guidelines go beyond systematic reviews by recommending evidence-based practices to manage patients. IOM produced a companion study about the standards for producing good guidelines including the role of scientifically valid systematic reviews.⁴⁵ To make practice guidelines actionable, many systems have incorporated decision support

systems and alerts that provide information about guideline-recommended actions at the point of care.

Electronic Health Records

When the American Recovery and Reinvestment Act of 2009 was passed with incentives for electronic health record (EHR) adoption, there was optimism that EHRs would contribute both to assessing quality performance and providing tools to ensure reliable delivery of high-quality care. The idea of seamless data flows that would help physicians and systems ensure they were delivering recommended care, offer shared decision making tools, and have real-time feedback on quality performance was appealing. Others have written eloquently about the failure of our current experience to meet the goal,⁴⁶ but it likely that we will redesign rather than eliminate EHRs from medical practice.

Quality Improvement Programs

Quality improvement programs have been organized at different levels in the health care system. The approaches include Plan-Do-Study-Act (PDSA) cycles, Lean, 6 Sigma, and total quality management. Typically, these approaches focus on changing a discrete practice (rather than larger systems changes) and are undertaken in local settings. Quality improvement also tends to be iterative and limited to the specific context in which the improvement strategy is undertaken, making systematic assessment of the impact of these programs challenging.

Putting It All Together: Are We Making Progress?

Over the last 50 years, we have become more aware that recommended care is not always offered or received, and we have implemented a variety of programs and supports to address that problem. We are measuring more dimensions of quality, more often, for more parts of the health care system. We are realigning incentives to reward better performance. We are trying to make it easier to know what works best for which patients. We are making that information more readily available at the point of care. There are individually focused and organizationally driven programs to improve quality. The public is more aware of variations in quality and some are more engaged in ensuring they receive the care they need or in choosing where to go for care based on

publicly available data. So, we must be doing better, right?

This is a hard question to answer. We have not had any large-scale, national assessment of quality since 2003. Much of the public reporting in quality focuses on subsets of the population defined by the setting of care (eg, hospitals), the payer (eg, Medicare, managed care), or a limited number of conditions (eg, heart attacks). We have some insights from AHRQ which publishes a regular report on quality and disparities.⁴⁷ The report includes a variety of process, outcome, and patient-experience measures from multiple federal data sources. The collection of indicators and report construction make it difficult to garner any overall sense of the current state of quality. Levine and colleagues⁴⁸ examined changes in a select number of process measures using data from the federal Medical Expenditures Panel Survey and found little change over the period from 2002 to 2013.

The National Committee for Quality Assurance (NCQA) has the longest standing public report on quality across multiple sectors. The current version of Healthcare Effectiveness Data and Information Set (HEDIS) includes 57 measures in multiple categories (prevention, chronic care management, overuse, patient experience) and is reported for commercial, Medicaid and Medicare enrollees. An examination of the 2018 State of Health Care Quality report underscores the variability in progress.⁴⁹ For example, breast and cervical cancer screening have shown little improvement, hovering around the low 70%*s*, whereas colorectal cancer screening among Medicare Advantage enrollees improved from 52.6% to 69.6% from 2004 to 2017. Blood pressure control among commercial Health Maintenance Organization (HMO) enrollees has improved from 39% to 62.2% from 1999 to 2017 and glycated hemoglobin (HbA1c) control has stayed about the same for both commercial and Medicare enrollees at around 60%.

The collection of studies on public reporting, pay for performance and value-based purchasing also lead to the conclusion that major improvements in quality have not occurred as a result of those initiatives. There have been pockets of improvement (“islands of excellence”) but few systematic, sustained, and substantial improvements.

Where Do We Go from Here?

Improving quality seems to be a daunting task. We may have been focused on the wrong approaches,

specifically, relying on payment incentives to create the conditions for improvement despite considerable research that finds little relationship between the way care is paid for, whether in the United States or other countries, and the appropriateness or quality of care delivered. It is not unreasonable to consider whether payment methods create barriers to improvement, but it is perhaps unreasonable to assume that changes in payment alone will address the problem. Another approach has been to focus on the measures including reducing the number of measurements,⁵⁰ aligning measures with other regulatory programs or developing different measures (generally under the heading of “measures that matter”). Most of these efforts are designed to reduce the burden of measurement (rather than improve quality) but this would require a level collaboration across state, federal, and private entities that has not yet emerged.

Improving quality requires a systems approach. This is particularly evident in primary care where the complexity of matching the unique characteristics of an individual patient to the wide array of necessary or recommended interventions in the context of a series of short, unplanned, and uncertainly scheduled visits exceeds the capacity of the human brain. Moreover, patients initiate much of their own care, move between coverage programs and providers, and have preferences that inform (rightfully) the care they receive—all which make delivery and measurement of quality care more complex. Health care systems that have made significant and sustained progress in quality have devoted time, people, and resources to creating systems—and still encounter challenges every day in managing improvement.⁵¹

The observation that quality is a systems issue leads some to wonder which components have the largest effect on performance—physicians, medical groups, health plans, patients, and other economic factors. One study found that the variance attributable to physicians ranged from 39% to 49%, compared with 13% to 24% for medical groups and 12% to 26% for the service area, depending on the measure.⁵² Studies using different measures, settings, and geographies come to different conclusions^{53–55} but all entities contribute to observed results.

An example from Kaiser Permanente illustrates the multiple components that may be required to achieve high performance. Between 2000 and 2013, Kaiser

Permanente–Northern California improved blood pressure control among patients with hypertension from 44% to 90%.^{56,57} The region undertook a coordinated effort with 6 critical components: leadership commitment, developing and building a hypertension registry, creating and updating an evidence-based guideline that included a drug treatment algorithm, routine feedback to medical centers and physicians on performance, medical assistant blood pressure visits without copays, and incorporating single-pill combination medication. This systems approach—with the work of everyone on the team aligned—has been replicated in other clinical areas with similar significant improvements in performance.

Because most physicians do not practice in organizations like these large systems, alternative mechanisms may be needed to create virtual or other systems that could support physicians on their quality journeys. This represents a potential role for Boards that are responsible for assessing physician quality through performance improvement modules that constitute a part of maintenance of certification. The ABFM, for example, could identify a comprehensive set of quality indicators that reflect excellence in primary care, develop systems to routinely assess physician performance on those indicators (including novel ways of extracting data), and provide feedback to physicians to facilitate improvement. They could certify programs or entities that provide good value in coaching physicians in quality improvement. Boards could also highlight where financial barriers exist to delivering quality care.

At the same time, Boards could invest in reimagining quality measurement, particularly for primary care physicians, and work with regulators and accreditors to provide safe harbors for those willing to be part of the needed innovations.^{58,59} There is considerable dissatisfaction with the current state of measurement and assessment but few serious efforts exist to make significant changes. This is a ripe area for physician leadership—and if physicians do not or cannot step up to lead these efforts, we may see more of the same ineffective efforts continue.

To see this article online, please go to: <http://jabfm.org/content/33/Supplement/S28.full>.

References

- Blumenthal D, Morone J. The heart of power: health and politics in the oval office. Berkeley, CA: University of California Press; 2009, 226.
- Morehead MA, Donaldson R, Sanderson RS, et al. A study of the quality of hospital care secured by a sample of teamster family members in New York City. New York, NY: Columbia University, School of Public Health and Administrative Medicine; 1964.
- Lyons TF, Payne BC. The Quality of Physicians' Health-Care Performance: A comparison against optimal criteria for treatment of the elderly and younger adults in community hospitals. *JAMA* 1974;227:925–8.
- Rhee SO. Factors determining the quality of physician performance in patient care. *Med Care* 1976;14:733–50.
- Wennberg J, Gittelsohn A. Small area variations in health care delivery. *Science* 1973;182:1102–8.
- Brook RH, Chassin MR, Fink A, Solomon DH, Koseoff J, Park RE. A method for the detailed assessment of the appropriateness of medical technologies. *Int J Technol Assess Health Care* 1986;2:53–63.
- Winslow CM, Koseoff JB, Chassin M, Kanouse DE, Brook RH. The appropriateness of performing coronary artery bypass surgery. *JAMA* 1988;260:505–9.
- Winslow CM, Solomon DH, Chassin MR, Koseoff J, Merrick NJ, Brook RH. The appropriateness of carotid endarterectomy. *N Engl J Med* 1988;318:721–7.
- Bernstein SJ, McGlynn EA, Siu AL, et al. The appropriateness of hysterectomy. A comparison of care in seven health plans. Health Maintenance Organization Quality of Care Consortium. *JAMA* 1993;269:2398–402.
- Hilborne LH, Leape LL, Bernstein SJ, et al. The appropriateness of use of percutaneous transluminal coronary angioplasty in New York State. *JAMA* 1993;269:761–5.
- Leape LL, Hilborne LH, Park RE, et al. The appropriateness of use of coronary artery bypass graft surgery in New York State. *JAMA* 1993;269:753–60.
- Bernstein SJ, Hilborne LH, Leape LL, Park RE, Brook RH. The appropriateness of use of cardiovascular procedures in women and men. *Arch Intern Med* 1994;154:2759–65.
- Kleinman LC, Koseoff J, Dubois RW, Brook RH. The medical appropriateness of tympanostomy tubes proposed for children younger than 16 years in the United States. *JAMA* 1994;271:1250–5.
- McGlynn EA, Brook RH. Ensuring quality of care. In Anderson RJ, Rice TH, Kominski GF, eds. *Changing the U.S. health care system: key issues in policy and management*. 2nd ed. San Francisco, CA: Jossey-Bass; 2001.
- Siu AL, Sonnenberg FA, Manning WG, et al. Inappropriate use of hospitals in a randomized trial of health insurance plans. *N Engl J Med* 1986;315:1259–66.
- Chassin MR, Koseoff J, Park RE, et al. Does inappropriate use explain geographic variations in the

- use of health care services? A study of three procedures. *JAMA* 1987;258:2533–7.
17. Leape LL, Park RE, Solomon DH, Chassin MR, Koseoff J, Brook RH. Does inappropriate use explain small-area variations in the use of health care services? *JAMA* 1990;263:669–72.
 18. Gray D, Hampton JR, Bernstein SJ, Koseoff J, Brook RH. Audit of coronary angiography and bypass surgery. *Lancet* 1990;335:1317–20.
 19. Pilpel D, Fraser GM, Koseoff J, Weitzman S, Brook RH. Regional differences in appropriateness of cholecystectomy in a prepaid health insurance system. *Public Health Rev* 1992–1993;20:61–74.
 20. McGlynn EA, Naylor CD, Anderson GM, et al. Comparison of the appropriateness of coronary angiography and coronary artery bypass graft surgery between Canada and New York State. *JAMA* 1994;272:934–40.
 21. Bernstein SJ, Brorsson B, Aberg T, Emanuelsson H, Brook RH, Werko L. Appropriateness of referral of coronary angiography patients in Sweden. *SECOR/SBU Project Group. Heart* 1999;81:470–7.
 22. Carlisle DM, Valdez RB, Shapiro MR, Brook RH. Geographic variation in rates of selected surgical procedures within Los Angeles County. *Health Serv Res* 1995;30:27–42.
 23. Schuster MA, McGlynn EA, Brook RH. How good is the quality of health care in the United States? *Millbank Q* 1998;76:517–63.
 24. Institute of Medicine (US) Committee on Quality of Health Care in America. Kohn LT, Corrigan JM, Donaldson MS, eds. *To err is human: building a safer health system*. Washington, DC: National Academies Press; 1999.
 25. Institute of Medicine (US) Committee on Quality of Health Care in America. *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: National Academies Press; 2001.
 26. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *N Engl J Med* 2003;348:2635–45.
 27. Kerr EA, McGlynn EA, Adams J, Keeseey J, Asch SM. Profiling the quality of care in 12 communities: results CQI study. *Health Aff (Millbank)* 2004;23:247–56.
 28. Asch SM, Kerr EA, Keeseey J, et al. Who is at greatest risk for receiving poor-quality health care? *N Engl J Med* 2006;354:1147–56.
 29. Mangione-Smith R, DeCristofaro AH, Setodji CM, et al. The quality of ambulatory care delivered to children in the United States. *N Engl J Med* 2007;357:1515–23.
 30. Hannan EL, Barach PR, Jacobs JP, Lipshultz SE, Laussen PC, eds. *Public reporting of cardiac data: pros, cons, and lessons for the future. Pediatric and Congenital Cardiac Care* 2015;1:467–77.
 31. Totten AM, Wagner J, Tiwari A, O'Haire C, Griffin J, Walker M. Closing the quality gap: revisiting the state of the science (vol. 5, public reporting as a quality strategy). *Evid Rep Technol Assess*, No 208.5. Rockville, MD: Agency for Healthcare Research and Quality; 2012.
 32. Campanella P, Vukovic V, Parente P, Sulejmani A, Ricciardi W, Specchia ML. The impact of public reporting on clinical outcomes: a systematic review and meta-analysis. *BMC Health Serv Res* 2016;16:296.
 33. Schlesinger M, Kanouse DE, Martino SC, Shaller D, Rybowski L. Complexity, public reporting, and choice of doctors: a look inside the blackest box of consumer behavior. *Med Care Res Rev* 2014;71:38S–64S.
 34. Werner R, Kolstad JT, Stuart KA, Polsy D. The effect of pay-for-performance in hospitals: lessons for quality improvement. *Health Aff (Millbank)* 2011;30:690–8.
 35. Mendelson A, Kondo K, Damberg C, et al. The effects of pay-for-performance programs on health, health care use, and processes of care: a systematic review. *Ann Intern Med* 2017;166:341–53.
 36. Lester H, Schmittiel J, Selby J, et al. The impact of removing financial incentives from clinical quality indicators: longitudinal analysis of four Kaiser Permanente indicators. *BMJ* 2010;340:c1898.
 37. Minchin M, Roland M, Richardson J, Rowark S, Guthrie B. Quality of care in the United Kingdom after removal of financial incentives. *N Engl J Med* 2018;379:948–57.
 38. Chee TT, Ryan AM, Wasfy JH, Borden WB. Current state of value-based purchasing programs. *Circulation* 2016;133:2197–205.
 39. Ryan AM, Krinsky S, Maurer KA, Dimick JB. Changes in hospital quality associated with hospital value-based purchasing. *N Engl J Med* 2017;376:2358–66.
 40. Medicare Payment Advisory Commission. *Report to the Congress: Medicare payment policy*. Washington, DC: MedPAC; 2019.
 41. Institute of Medicine. *Finding what works in health care: standards for systematic reviews*. Washington, DC: The National Academies Press; 2011.
 42. Peterson LE, Blackburn BE, Puffer JC, Phillips RC. Family physicians' quality interventions and performance improvement through the ABFM diabetes performance in practice module. *Ann Fam Med* 2014;12:17–20.
 43. Agency for Healthcare Research and Quality. *About the evidence-based practice centers program*. Rockville, MD. 2015. Available from: <https://www.ahrq.gov/research/findings/evidence-based-reports/overview/index.html>. Accessed October 27, 2019.
 44. Cochrane Collaboration. *Cochrane Database of Systematic Reviews*. Cochrane Library. 2019.

- Available from: <https://www.cochranelibrary.com/cdsr/about-cdsr>. Accessed October 27, 2019.
45. Graham R, Mancher M, Miller Wolman D, Greenfield S, Steinberg E, eds. *Clinical practice guidelines we can trust*. Washington, DC: The National Academies Press; 2011.
 46. Wachter R. *The digital doctor: hope, hype, and harm at the dawn of medicine's computer age*. McGraw-Hill Education; 2015.
 47. Agency for Healthcare Research and Quality. 2018. *National Healthcare Quality and Disparities Report*. 2019. Available from: <https://www.ahrq.gov/research/findings/nhqrdr/nhqrdr18/index.html>. Accessed October 27, 2019.
 48. Levine DM, Linder JA, Landon BE. The quality of outpatient care delivered to adults in the United States, 2002 to 2013. *JAMA Intern Med* 2016;176:1778–90.
 49. National Committee for Quality Assurance. *State of health care quality, 2018*. 2019. Available from: <https://www.ncqa.org/report-cards/health-plans/state-of-health-care-quality-report/>. Accessed October 2, 2019.
 50. Institute of Medicine. *Vital signs: core metrics for health and health care progress*. Washington, DC: The National Academies Press; 2015.
 51. Schilling L, Chase A, Liu AY, et al. Kaiser Permanente's performance improvement system, part 1: from benchmarking to executing on strategic priorities. *Jt Comm J Qual Saf* 2010;36:484–98.
 52. Rodriguez HP, Scoggins JF, von Glahn T, Zaslavsky AM, Safran DG. Attributing sources of variation in patients' experiences of ambulatory care. *Med Care* 2009;47:835–41.
 53. O'Connor PJ, Rush WA, Davidson G, et al. Variation in quality of diabetes care at the levels of patient, physician, and clinic. *Prev Chronic Dis* 2008;5:A15.
 54. Selby JV, Schmittiel JA, Lee J, et al. Meaningful variation in performance: what does variation in quality tell us about improving quality? *Med Care* 2010;48:133–9.
 55. Zaslavsky AM. Using hierarchical models to attribute sources of variation in consumer assessments of health care. *Statist Med* 2007;26:1885–900.
 56. Jaffe MG, Lee GA, Young JD, Sidney S, Go AS. Improved blood pressure control associated with a large-scale hypertension program. *JAMA* 2013;310:699–705.
 57. Jaffe MG, Young JD. The Kaiser Permanente Northern California story: improving hypertension control from 44% to 90% in 13 years (2000–2013). *J Clin Hypertens* 2016;18:260–1.
 58. McGlynn EA, Schneider EC, Kerr EA. Reimagining quality measurement. *N Engl J Med* 2014;371:2150–3.
 59. McGlynn EA, Kerr EA. Creating safe harbors for quality measurement innovation and improvement. *JAMA* 2016;315:129–30.