Robert Wood Johnson Foundation

Published: September 4, 2012
Program Results Report
Grant ID: 68275 in QE 3

How Physicians Perform

Building empirical evidence for scoring and classifying physician performance

SUMMARY

During 2010–2012, researchers for the Pacific Business Group on Health explored evidence-based models for scoring physicians on the health care they provide and analyzed mechanisms by which physicians could correct errors in data on their performance evaluations.

The models took into consideration that physicians' practices are not all alike. Physicians have differing caseloads and patient populations and practice in different types of groups. These and other factors can influence performance.

The project was funded under a 2010 solicitation, Improving Quality and Value in Health Care: Ideas from the Field, from the Quality/Equality Program Management Team at RWJF. This solicitation sought to promote learning and knowledge about innovative efforts that address health care quality and value problems, by studying such efforts in the following specific areas, in order to understand how they may lead to better health care quality and lower costs:

- Value-based purchasing
- Data collection and aggregation for performance measurement
- Quality improvement support
- Public reporting of provider performance.

Read the Introduction for more information.

The Pacific Business Group on Health, founded in 1989 and based in San Francisco, is a coalition of 50 major health care purchasers active in California. It works with health insurance plans, physician groups, consumer organizations, and other stakeholders to improve the quality and affordability of health care.

Key Findings

In reports to the Robert Wood Johnson Foundation (RWJF), journal articles and an interview for this report, project staff stated that:

- Of three statistical models evaluated for rating physician performances, the "adjusted opportunities model" was the "the most appealing option."
- Combining data on a physician's group and/or practice yielded a more accurate reflection of physician performance than data on the physician alone.
- When physicians are asked to correct data on their performance evaluations, they are unlikely to so.

Funding

RWJF supported this project from November 2010 through January 2012 with a grant of $200,009. It was funded through Improving Quality and Value in Health Care, a 2010 solicitation, and was one of 12 funded projects focused on how to achieve better health care quality and lower costs.

CONTEXT

Ninety-six percent of commercial health plans had physician quality incentives in operation or under development, a recent survey found.[1] Yet, according to several research articles, opposition to individual-level profiling has been strong within the physician community, particularly regarding initiatives to rank and publicly report performance.[2]

Methods to classify performance must be rooted in strong empirical evidence in order to gain full acceptance from physicians and providers. That evidence did not exist at the start of this project, according to the Pacific Business Group on Health in its proposal to RWJF.

In addition, the proposal stated:

- "Differences in the case mix of individual practices lead physicians to review rating systems suspiciously."

---

[1] "2010 National P4P Survey," Med-Vantage, Inc. Available online.

[2] Three articles make this point: (1) Werner RM and Asch DA. "The Unintended Consequences of Publicly Reporting Quality Information." *Journal of the American Medical Association*, 293(10): 1239–1244, 2005; abstract available online. (2) Wharam JF, Paasche-Orlow MK, Farber NJ, et al. "High Quality Care and Ethical Pay-for-Performance: A Society of General Internal Medicine Policy Analysis." *Journal of General Internal Medicine*, 249(7): 854–859, 2009; available online and (3) Rosenbaum S, Kornblet S, and Borzi PC. "An Assessment of Legal Issues Raised in 'High Performing' Health Plan Quality and Efficiency Tiering Arrangements: Can the Patient be Saved?" *Health Care Policy Reporter*. Arlington, VA: Bureau of National Affairs; 2007; available online.

- "Commercial insurance program populations are often too small to produce reliable and meaningful quality performance results at the individual physician level."
- "Physician quality transparency initiatives must include processes for physicians to review their performance results and make corrections."

RWJF's Interest in This Area

The 2010 solicitation Improving Quality and Value in Health Care sought to promote learning and knowledge about innovative efforts that addressed health care quality and value problems. Heightened concern across the country about the ballooning costs and poor quality of health care led to the funding.

The 12 grants, totaling $3.1 million, addressed issues in value-based purchasing, quality improvement support, public reporting of provider performance, and data collection and aggregation for performance measurement.

Among the initiatives RWJF had previously funded in the area of performance measurement were:

- The High Value Health Care Project, a $15.8 million initiative that ran from 2007 to 2010, aimed to build the initial infrastructure for a nationwide performance measurement and reporting system. The project included a component about evaluating physician performance. See the Program Results for more information.
- In 2005–2007, the National Committee for Quality Assurance examined applications of pay-for-performance in behavioral health care.[3]
- In 2009, Cornell University, the Joan and Stanford I Weill Medical College studied whether medical practices in less affluent areas score lower on quality measures and how pay-for-performance might improve performance scores.[4]

Since 2004, RWJF has made numerous grants to the Pacific Business Group on Health to support projects in quality improvement, accountability, and transparency in health care.[5]

THE PROJECT

From November 2010 through January 2012, researchers for the project explored evidence-based models for scoring physicians on the health care they provide.

---

[3] This project, funded by Grant ID#s 51650 and 55816 was part of a national program, *Depression in Primary Care.* See Program Results.

[4] The project was funded by grant ID# 65453 under a solicitation, Examining the Effects of Public Reporting and Pay-for-Performance on Health Care Quality.

[5] Grant ID#s 50791 and 52352 focused on the value of patient-reported quality information for quality improvement and consumer choice. ID#s 56186, 58867, 67565, and 68765 were for the Consumer-Purchaser Disclosure Project.

The models took into consideration that physicians' practices have differing caseloads and patient populations and that physicians practice in different types of groups because these and other factors can influence performance. The researchers also analyzed mechanisms by which physicians could correct data on their performance evaluations.

After conducting a literature review and preparing a data analysis plan, researchers obtained medical and pharmacy claims information, filed between October 1, 2007, and September 30, 2008, from three California commercial health insurance plans.[6]

The data, which did not identify patients by name, had been submitted to the California Physician Performance Initiative.[7] The data included patient zip codes in addition to age and gender, which allowed researchers to make certain assumptions about a patient's level of education, income, ethnicity, and language.

Staff then developed a dataset for results on four evidence-based measures for diabetes screening and treatment:[8]

- Patient received an LDL-cholesterol screening test during the measurement year.
- Patient received an HbA1c (blood sugar) test during the measurement year.
- Patient had a screening for nephropathy (kidney disease) during the measurement year or was observed with nephropathy symptoms.
- Patients—not all of them diabetic—who had been prescribed at least a 180-day supply of one or more of the following categories of drugs received at least one monitoring session:
  - ACE inhibitors or ARBs for hypertension
  - Digoxin for congestive heart failure
  - Diuretics[9]

The dataset comprised 1,418 primary care physicians in 1,058 practices within 133 groups. All told, they cared for 36,889 patients ages 18 to 75.

---

[6] Anthem Blue Cross of California, Blue Shield of California, and United Healthcare.

[7] The initiative, started in 2006, measures and reports the quality of patient care provided by individual physicians in California. It is a project of the California Cooperative Healthcare Reporting Initiative, a statewide collaborative of physician organizations, health plans, purchasers, and consumers convened in 1993 by the Pacific Business Group on Health "to help consumers and purchasers make informed health care purchasing decisions." See *www.cchri.org* and *www.cchri.org/programs/programs_CPPI.html*.

[8] The measures are standards recommended in the Healthcare Effectiveness Data and Information Set (HEDIS) of the National Committee for Quality Assurance.

[9] Patients had at least one therapeutic monitoring event for the therapeutic agent in the measurement year.

Researchers blended results from all four measures, according to Kristi Alvarez, manager of health performance information for the Pacific Business Group on Health. Researchers said this gave a more accurate picture than depending on any one measure.

For example, Alvarez said, if a physician gives one of the recommended screening tests to all 10 diabetic patients he or she sees, it's reasonable to assume he or she will do this screening for the 11th. If the physician has only done it for two patients, however, one cannot assume that he or she will do it for a third.

The researchers also factored in patient demographics. For example, patients with lower socioeconomic status often do not get the recommended treatment for various reasons. The researchers took this into consideration when rating physicians who had many such patients.

Statistical Models

Researchers selected three statistical methods to get a composite score for each physician based on all four measures:

- A "simple opportunities model" is based on the number of opportunities a physician had to administer the prescribed treatment or screening and the number of times he or she did so. This score is then adjusted to allow for patient demographic characteristics and the effect of the physician's practice and group.
- The PRIDIT model converts physician scores to percentiles and then ranks the physicians; it can also be used for hospitals.
- The Item Response Theory model, a variation on the PRIDIT model

Researchers evaluated all three models for reliability, transparency, and simplicity of computing scores.

The Process for Correcting Errors in Performance Data

"A major concern about physician quality measurement is the possibility that the data used to score the measures are incomplete or incorrect," project director Ted von Glahn, MSPH, noted. "This concern is heightened when quality scores rely on patient administrative data, which is primarily used for billing and payment."

The California Physician Performance Initiative had sent participating physicians materials offering them the chance to make corrections to their data.

Project researchers assessed responses across a dataset comprising 12,749 physicians and 17 treatment measures from October 1, 2007, to September 30, 2008. The dataset represented some one-third of the commercial market in California. Researchers assessed:

- The percentage of physicians who requested corrections materials and those who submitted them
- The frequency and type of corrections
- Differences between correcting and noncorrecting physicians, including patient demographics
- The impact of corrections on overall scores

Research Papers

Project Director von Glahn; Principal Investigator William H. Rogers, PhD, of the Institute for Clinical Research and Health Policy Studies at Tufts Medical Center in Boston; and Alvarez produced three articles that are under review at peer-reviewed journals—see the Bibliography.

FINDINGS

In reports to RWJF, the research articles and an interview, project staff stated that:

- **The adjusted "simple opportunities model"[10] was "the most appealing option" to rate physician performances.**

  Researchers found this model more transparent and easier to compute than the other two models. The project's physician advisory group, made up of medical directors from the participating health plans, and a subset of medical groups in California, approved their recommendation.

- **Combining data on a physician's group and/or practice yielded a more accurate reflection of physician performance than data on the physician alone.**

  Group or practice characteristics that influence physician performance include: systems used for monitoring claims and keeping medical records, protocols for following up with patients, cross-covering, and staff routines.

  Blending group or practice data brought the proportion of reliably classified physicians to 80 percent—or 65 percent of those surveyed, depending on the method of computation. Eighty percent is a good percentage for a rating system, Alvarez said, and "is going to get you a lot of attention."

  The physician advisory group, however, expressed some concern about how blending affected the evaluations of physicians in solo practice and about the fact that some medical groups are more cohesive than others.

[10] The simple opportunities model is adjusted to allow for patient demographic characteristics and the effect of the physician's practice and group.

- **When physicians are asked to correct data on their performance evaluations, they are unlikely to do so.**

  Of the 12,749 physicians in the sample, 15 percent requested corrections materials and just over half of them, 8.3 percent, completed the forms. There was no significant difference in the performance rate of physicians who corrected their data and those who didn't, but quality scores increased by 9.7 percent for correcting physicians.

  Physicians affiliated with medical groups were more likely to correct data, presumably because they had more staff support. This was one reason why researchers attributed the low response rate to physicians finding the process "burdensome."

  One recurring issue, Alvarez said, is whether a physician should be able to remove a patient from the sample for rating purposes if the patient refuses a recommended treatment. This accounted for the largest share of corrections.

  "It is the shared responsibility of doctor and patient to get the right care," she said. "A doctor's communication and other skills are a factor in the percentage of patients who are screened or treated, but how that's applied would be a policy decision that's up to the rating agency."

  "In a pay-for-performance application, putting doctors who do not correct at a disadvantage in terms of quality scores may be fair," the researchers said, but "generating biased scores for doctors who do not correct is a disservice to the public."

Limitations

In addition to the relatively limited universe from which the 12,000 physicians were pulled (just three health plans), "one recurring limitation includes working with administrative claims data for purposes other than billing," Alvarez said.

Another challenge is devising performance measures that would be easily understood by physicians, patients, and providers, but, said Alvarez, "the complexity of statistical methods required often creates a barrier that is difficult to overcome."

LESSONS LEARNED

1. **The adjusted opportunities model should be studied with measures other than those for the treatment of diabetes.** Expanding the process to areas such as cardiovascular, respiratory, and preventive care "would add to the evidence and our understanding of the method's widespread feasibility." (Project Director/von Glahn)

2. **The physician corrections process needs improvement.** One possibility: A secure, Web-based physician portal to "reduce the burden of corrections while continuing to address physicians' concerns." A larger database including all claims payers would also produce more accurate results. (Project Director/von Glahn)

3. **Improved data collection on physician and medical group and practice affiliations is needed.** "We believe that this infrastructure will be a required component of the Medicare and commercial value-based performance initiatives." (Project Director/von Glahn)

AFTERWARD

Researchers presented the project's findings in a webinar entitled "Physician Feedback and Value-Based Modifier Programs" on February 29, 2012. More than 500 providers nationwide attended through the Center for Medicare & Medicaid Service's Special National Provider Call Series.

Prepared by: Paul Jablow
Reviewed by: Kelsey Menehan and Molly McKaughan
Program Officer: Claire B. Gibbons
Program Area: Quality/Equality
Grant ID# 68275 in QE 3 Solicitation
Project Director: Ted von Glahn (415) 615-6318; tvonglahn@pbgh.org

---

BIBLIOGRAPHY

Articles

Rogers WH, Alvarez K and von Glahn T. "Models for Evaluating Physician Quality: A Statistical Comparison." Unpublished.

Rogers WH, Alvarez K and von Glahn T. "Bias in Physician Corrections of Claims-Based Performance Measures." Unpublished.

Rogers WH, Alvarez K and von Glahn T. "Blending Group and Practice Site Scores to Increase the Reliability of Physician Quality Information." Unpublished.