CENTER for STUDYING HEALTH SYSTEM CHANGE

Issue Brief

Findings from HSC

NO. 110 • MARCH 2007

CONSUMER TOLERANCE FOR INACCURACY IN PHYSICIAN PERFORMANCE RATINGS: ONE SIZE FITS NONE

By Matthew M. Davis, Judith H. Hibbard and Arnold Milstein

Health plans increasingly use physician performance ratings, but some physicians are concerned that measurement inaccuracies may jeopardize their reputations and livelihoods. Absent from the debate thus far are consumer views about how accurate physician ratings need to be for various uses. Consumer tolerance for inaccuracy in physician performance ratings varies widely, according to a new national study by the Center for Studying Health System Change (HSC). At least one-third of adults have a low tolerance for inaccuracy (5 percent or less), but more than one of every five adults would tolerate ratings that were 20 percent-50 percent inaccurate. Consumers' relatively higher tolerance for inaccuracy when used for public reporting and tiered networks may speed these uses of physician performance ratings by health plans. However, consumers' lower tolerance for inaccurate ratings when choosing their own physicians and paying physicians for performance may hinder such uses.

Physician Performance Ratings—How Tolerant Are Consumers of Inaccuracy?

As the health care system continues to expand measurement of quality and cost-efficiency, individual physician performance ratings are gaining attention as a way to improve physician practice and link patients with better-performing clinicians.[1] Because of limitations in current data sources and measures for rating physicians, performance ratings inevitably incorporate imprecision.

Not surprisingly, some physicians are uncomfortable with payers' use of imprecise ratings, because they believe their reputations and livelihood might be unfairly jeopardized.[2] On the other hand, health plans, purchasers and consumer leaders point out that public ratings, even if somewhat inaccurate, may stimulate more improvement than a performance-blind environment without ratings.[3,4]

Absent from this debate has been the voice of consumers, whose clinical care and health care-seeking behavior may be influenced if physician performance ratings are widely adopted. Consumer tolerance for inaccuracy in physician performance ratings varies widely, according to an HSC December 2006 national survey designed to explore public tolerance for inaccuracy in physician performance ratings and to examine factors associated with tolerance levels (see Data Source).

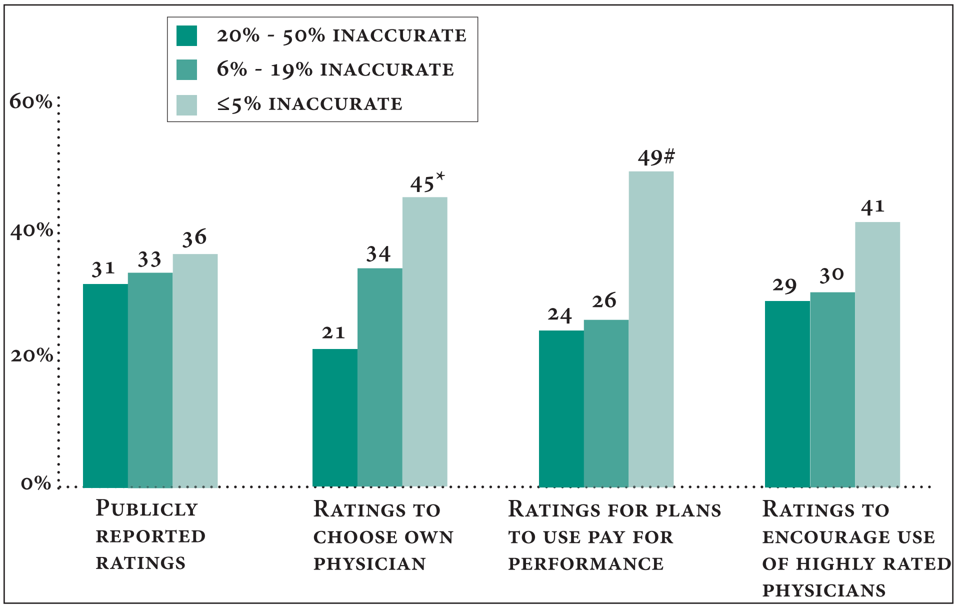
The survey measured consumer acceptance for measurement error in physician performance ratings for four purposes in which such ratings might be used: releasing ratings to the general public, using ratings to choose one's own primary care physician, using ratings to vary physicians' payment rates based on performance, and using ratings to encourage consumers to seek care from more highly rated physicians via tiered-benefit plans.

As part of the survey, inaccuracy was described for respondents in the context of performance ratings that would classify physicians as higher performing or lower performing. Respondents were told that ratings that were 80 percent accurate/20 percent inaccurate would incorrectly classify 20 of every 100 physicians. For example, 20 percent might be classified as higher-performing physicians when they are actually lower-performing physicians or vice-versa. How the question was framed—either as misclassification of higher-performing physicians or lower-performing physicians—did not influence respondents' preferences. Respondents were offered choices of rating inaccuracy ranging from 1 percent or less (i.e., at least 99 percent accurate) to as much as 50 percent (at least 50 percent accurate).

Regardless of the use of physician performance ratings, consumers reported a wide range of tolerance for rating inaccuracy (see Figure 1). The most common response for each application of physician performance ratings was low tolerance for inaccuracy (5 percent or less), and for three of the four uses at least 40 percent of consumers had low tolerance for inaccuracy.

Figure 1

Consumer Tolerance for Inaccuracy of Physician Performance Ratings, by Rating Application



* Difference from public ratings statistically significant at $p < .05$.
# Difference from public ratings and for ratings to encourage use of highly rated physicians statistically significant at $p < .05$.
Source: Center for Studying Health System Change, national survey conducted by Knowledge Networks, Inc., December 2006

On the other hand, more than 20 percent of consumers were comfortable with inaccuracy of 20 percent or more (high tolerance) across all four uses.

Consumers were relatively tolerant of inaccuracy when ratings were used for public reporting and tiered networks. Consumers demonstrated the lowest tolerance for inaccuracy in two circumstances—using ratings to choose their own physicians and insurance plans paying physicians differently based on ratings. Consumers' low tolerance for inaccuracy when choosing their own physician is expected. Consumers' even lower tolerance for inaccuracy in the use of ratings by plans to implement pay-for-performance programs may indicate consumers' concerns about the use of physician payment incentives in general.

Consumers were quite consistent in their tolerance for inaccuracy across the different potential uses of physician ratings. For example, among those who had the lowest tolerance for inaccuracy regarding choice of their own physician, 76 percent also had the lowest tolerance for inaccuracy regarding pay-for-performance applications.

Publicly Available Information and Choosing a Physician

People who tolerated the lowest levels of inaccuracy in physician performance ratings for purposes of public information and for choosing a physician for themselves were disproportionately middle-aged (45-64 years) and likely to have a regular doctor.

Respondents were asked about the extent to which they believe that physicians differ in following quality-of-care guidelines and how they would rate the importance of physician performance characteristics such as on-time care and preventing treatment complications. Regarding release of physician performance ratings to the general public, there were no associations between these attitudes and respondents' tolerance for inaccuracy.

In contrast, when it came to choosing one's own physician based on performance ratings, the tolerance for inaccuracy was lowest among those who believe that physicians do not differ in following quality-of-care guidelines (see Table 1); this suggests that consumers are looking to ratings to help them distinguish physicians when they cannot themselves. Tolerance for inaccuracy was also lowest among those who believe that on-time care and preventing treatment complications are very important in judging physician performance, and they believe these are key factors for plans to consider as they compose ratings measures.

Consumers' tolerance for rating inaccuracy in release of public ratings and in choosing a physician for themselves did not differ with respect to prior use of consumer ratings, perceived helpfulness of consumer ratings, presence of a chronic health condition, type of insurance coverage, how involved consumers are in managing their own health care, their attitudes about the extent to which physicians differ with regard to other performance measures (bedside manner, cost-effective care or preventing complications), perceived importance of other measures in determining physician performance (bedside manner, quality, cost-effectiveness), or respondent characteristics such as education, income, race/ethnicity and gender.

Pay-for-Performance Programs

Similar to the use of ratings for selecting their own physicians, people who expressed a low tolerance for inaccuracy in physician performance ratings for purposes of pay-for-performance initiatives by health plans were more likely to have the attitude that physicians do not differ in following quality-of-care guidelines (see Supplementary Table 1). People with low tolerance for inaccuracy also believed that physicians' ability to prevent complications should be a very important consideration in judging physician performance. Otherwise, low tolerance for inaccuracy regarding pay-for-performance was not associated with any health or sociodemographic characteristics, with use of consumer ratings in general, or with other attitudes about how physicians perform or on what basis their performance should be judged.

Encouraging Care from Highly Rated Physicians

People who had the lowest tolerance for inaccuracy in the use of physician performance ratings to encourage care from highly rated physicians (e.g., in tiered-health plans) exhibited a different set of characteristics than for the other potential ratings uses. In this case, people with the lowest tolerance for inaccuracy were significantly

Table 1

Choosing One's Own Physician—Respondent Characteristics, Beliefs and Attitudes and Tolerance for Inaccuracy in Physician Performance Ratings

Characteristics, Beliefs and Attitudes	High Tolerance for Inaccuracy (20% - 50%)	Moderate Tolerance for Inaccuracy (6% - 19%)	Low Tolerance for Inaccuracy (≤5%)
Overall	21%	34%	45%
Age			
18-44	24	36	40
45-64	18	29	53
65+	19	39	43
Have a Regular Doctor			
Yes	19	34	47
No	28	34	39
"Regarding following guidelines for quality of care (giving appropriate care), physicians are..."			
Usually the same	19	25	56
Usually a little bit different	20	39	41
Usually a lot different	26	32	42
"How important in judging the performance of a doctor is seeing patients on time?"			
Very important	16	32	53
Somewhat important	22	35	44
Not important	25	62	14
"How important in judging the performance of a doctor is preventing complications from treatment?"			
Very important	17	34	49
Somewhat important	36	39	35
Not important	74	26	0

Notes: Consumer tolerance for rating inaccuracy in choosing their own physician did not differ by other respondent characteristics, including education, income, insurance coverage, race/ethnicity and gender. Rows may not sum to 100% due to rounding. All differences across levels of tolerance for inaccuracy were statistically significant at p<.05.

Source: Center for Studying Health System Change, national survey conducted by Knowledge Networks, Inc., December 2006

more likely to have private insurance vs. other types of coverage and more likely to have excellent self-reported health status vs. poorer health status. They also were more likely to say that bedside manner should not be an important consideration in judging physician performance (data not shown). Otherwise, low tolerance for inaccuracy in the use of ratings to encourage care from highly rated physicians was not associated with the other health and sociodemographic characteristics measured, with use of consumer ratings, or with other attitudes about physician performance and measurement.

Experience with Ratings

Use of physician performance ratings by consumers may correspond to their general use of ratings for goods and services. Respondents indicated on this survey whether they had used such ratings in the past, and if they had, they were asked to indicate how helpful they had found general ratings of consumer goods and services (e.g., *Consumer Reports*) or sources about doctors (e.g., local magazines, or Web sites such as *Healthgrades.com* or *rateyourmd.com*).

Overall, 45 percent had used consumer ratings in the past. People with higher levels of education and those with higher incomes were significantly more likely to have used consumer ratings than other groups. Non-Hispanic blacks were significantly less likely to have used such ratings than non-Hispanic whites, Hispanics and other non-Hispanic respondents. Individuals with chronic conditions had used consumer ratings less than individuals without chronic conditions. Patterns of ratings use otherwise did not differ by age, gender or whether someone had a regular doctor.

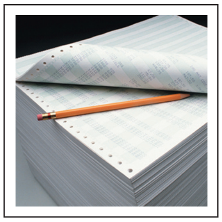
Of those who had ever used consumer ratings, 93 percent had used general ratings sources and 50 percent had used ratings of physicians. Information from general ratings sources was rated as helpful by 44 percent of those who had used such sources, compared with only 13 percent who found that ratings sources about doctors were helpful.

Implications

Consumer tolerance for inaccuracy in physician performance ratings is remarkably varied. At least one-third prefer that the ratings be no more than 5 percent inaccurate, while more than one-fifth of consumers tolerate rating inaccuracy of 20 percent or more, depending on the use of the ratings. These findings have several implications for health plans' efforts to implement physician performance ratings with currently available measures.

Consumer tolerance of inaccuracy in physician performance ratings is likely much higher than physicians' tolerance. It is possible that consumers are more tolerant of inaccuracy because they see flawed information as preferable to no information at all. Physicians, meanwhile, see the risks to their practice and reputation. They also are more aware of the technical aspects of performance and the considerable challenges of performance measurement.[5]

Major limitations in readily available clinical data sources and resulting performance measures make it likely that error in most individual physician ratings exceeds 5 percent. This raises a question: Should plans publicly report the level of rating inaccuracy? Consumers could then decline to use ratings that they believe are not accurate enough. In declining to use ratings, consumers run



Data Source

This Issue Brief presents findings from a nationally representative household survey conducted by Knowledge Networks, Inc., for the Center for Studying Health System Change. The survey was administered from Dec. 14-18, 2006, to a randomly selected group of adults aged 18 years and older on the Knowledge Networks standing panel, which was compiled through random-digit-dialed methods to create a panel that closely resembles the U.S. population and has access to the Internet (if households do not have such access, Knowledge Networks provides free access). The sample of 1,057 people was stratified so that half of the respondents reported chronic doctor-diagnosed conditions (e.g., heart disease, asthma, diabetes). The respondent group was subsequently sample-weighted so that the findings are representative of the U.S. population. Among Knowledge Networks panel members contacted to participate in the survey, the response rate was 64 percent. All results are presented in their weighted form to be nationally representative.

ISSUE BRIEFS are published by the Center for Studying Health System Change.

600 Maryland Avenue, SW, Suite 550
Washington, DC 20024-2512
Tel: (202) 484-5261
Fax: (202) 484-9258
www.hschange.org

President: Paul B. Ginsburg

risks such as choosing physicians whose performance is indeed worse than peer providers' or forgoing financial rewards from their plans when they seek care outside the group of highly rated physicians. Moreover, consumers who believe there are minimal differences in physician performance are likely less interested in performance data. If a substantial proportion of consumers decline to respond to ratings, plans and provider leaders will need to enhance the accuracy of their ratings, or consider other means of trying to improve care.

Communicating the accuracy of ratings, along with all the complexities associated with interpreting the meaning of the ratings, will substantially increase the complexity for consumers. Since complexity is already high and has been shown to be a barrier to consumers using public reports,[6] adding another layer of complexity may not be the best approach.

In the short term, health plans will be challenged to make ratings more accurate. Currently, ratings are based chiefly on analyses of single insurer's claims and enrollment data, which have acknowledged limitations in measurement reliability. Broader implementation of electronic medical records that are expressly designed to generate more robust performance measures will help considerably with ratings accuracy, but this development is not imminent. Health plans that encourage physicians to self-report supplementary clinical data may be able to push their own ratings efforts ahead faster, but this is usually accomplished with physician incentives that add to the cost of health insurance, and physician self-reports are rarely audited for accuracy.

Plans may incorporate patients' assessments of physician performance, such as those included on patient-driven Web sites. These may include patient-experience reports (e.g., physicians' bedside manner) and patient-observable quality-of-care events, such as checking the feet of diabetic patients. Plans also may seek to pool claims data with other private plans or, if it were permitted, with Medicare to achieve sample sizes needed for greater reliability in physician performance measurement. Finally, some physicians suggest limiting measurement and rating to multi-physician groups, which reduces some problems related to reliability of measurements but creates validity problems because between-physician performance variation can be significant within physician groups.

In an environment with broad variation in consumer tolerance for ratings inaccuracy, the most feasible near-term compromise may be for payers to convey the level of inaccuracy associated with their physician rating method. Questions remain about how to convey inaccuracy meaningfully and clearly. This study found that it did not matter to respondents whether inaccuracy was framed as a problem of misclassifying higher-performing or lower-performing doctors. However, little is known about consumers' expectations for the accuracy of ratings outside the health care sector or how best to explain measurement error to consumers in an understandable way. Implementation of performance ratings for physicians may compel further research to answer these questions, as health plans seek to encourage faster physician performance improvement, physicians push to address imperfect ratings, and consumers and plan sponsors are left wondering where to find the best value for their health care spending. ■

Notes

1. Department of Health and Human Services, "CMS News: Medicare to Provide Beneficiaries with Information on Physician Performance" (Feb. 15, 2007). Available at *http://www.cms.hhs.gov/apps/media/press_releases.asp*. Accessed March 19, 2007.

2. American Medical Association, "Regence Calls Off Flawed Physician Profiling." Available at *http://www.ama-assn.org/ama/pub/category/17128.html*. Accessed March 19, 2007.

3. Hibbard, Judith H., Jean Stockard and Martin Tusler, "Does Making Hospital Performance Public Increase Quality Improvement Efforts?" *Health Affairs,* Vol. 22, No. 2 (March/April 2003).

4. Hibbard, Judith H., Jean Stockard and Martin Tusler, "The Long-Term Effect of Public Performance Reporting on Hospital Quality Improvement, Market Share, and Reputation: Evidence from a Controlled Experiment," *Health Affairs,* No. 24, Vol. 4 (July/August 2005).

5. Pham, Hoangmai H., et al., "Care Patterns in Medicare and Their Implications for Pay for Performance, *New England Journal of Medicine*, Vol. 356, No. 11 (March 15, 2007).

6. Hibbard, Judith H., et al., "Strategies for Reporting Health Plan Performance Information to Consumers: Evidence from Controlled Studies," *Health Services Research*, Vol. 37, No. 2 (April 2002).

Consumer Tolerance for Inacurracy in Physician Performance Ratings: One Size Fits None

Supplementary Table

Supplementary Table 1

Pay-for-Performance Programs—Respondent Beliefs and Attitudes and Tolerance for Inaccuracy in Physician Performance Ratings

Resondent Beliefs and Attitudes	High Tolerance for Inaccuracy (20% - 50%)	Moderate Tolerance for Inaccuracy (6% - 19%)	Low Tolerance for Inaccuracy (≤5%)
Overall	24%	26%	49%
"Regarding following guidelines for quality of care (giving appropriate care), physicians are..."			
Usually the same	20	19	61
Usually a little bit different	26	30	44
Usually a lot different	25	26	51
"How important in judging the performance of a doctor is preventing complications from treatment?"			
Very important	21	27	52
Somewhat important	45	25	31
Not important	74	0	26

Note: Rows may not sum to 100% due to rounding.

All differences across levels of tolerance for inaccuracy were statistically significant at $p<.05$.

Source: Center for Studying Health System Change, national survey conducted by Knowledge Networks, Inc., December 2006