

# Relationship Between Patient Panel Characteristics and Primary Care Physician Clinical Performance Rankings

Clemens S. Hong, MD, MPH

Steven J. Atlas, MD, MPH

Yuchiao Chang, PhD

S. V. Subramanian, PhD

Jeffrey M. Ashburner, MPH

Michael J. Barry, MD

Richard W. Grant, MD, MPH

PHYSICIANS HAVE INCREASINGLY become the focus of quality performance measurement. Many health care systems now use physician clinical performance assessment as part of their re-credentialing process—to support health care choices for consumers and to guide quality improvement and cost-containment efforts. Partly in response to significant variation in the quality of care,<sup>1-3</sup> pay-for-performance and public reporting programs have become widely adopted approaches to influence clinician performance.<sup>4</sup> These programs use performance incentives including cash payments and public reports to motivate clinicians, practice groups, and health care systems to achieve specific health care quality goals.<sup>5-8</sup>

An intrinsic assumption underlying physician clinical performance assessment is that the measures represent physician performance. However, the same physician may have higher or lower measured quality scores depending on the panel of patients he or she manages. The association of patient panel characteristics with physician quality scores could lead to inaccurate physician clinical per-

**Context** Physicians have increasingly become the focus of clinical performance measurement.

**Objective** To investigate the relationship between patient panel characteristics and relative physician clinical performance rankings within a large academic primary care network.

**Design, Setting, and Participants** Cohort study using data from 125 303 adult patients who had visited any of the 9 hospital-affiliated practices or 4 community health centers between January 1, 2003, and December 31, 2005, (162 primary care physicians in 1 physician organization linked by a common electronic medical record system in Eastern Massachusetts) to determine changes in physician quality ranking based on an aggregate of Health Plan Employer and Data Information Set (HEDIS) measures after adjusting for practice site, visit frequency, and patient panel characteristics.

**Main Outcome Measures** Composite physician clinical performance score based on 9 HEDIS quality measures (reported by percentile, with lower scores indicating higher quality).

**Results** Patients of primary care physicians in the top quality performance tertile compared with patients of primary care physicians in the bottom quality tertile were older (51.1 years [95% confidence interval {CI}, 49.6-52.6 years] vs 46.6 years [95% CI, 43.8-49.5 years], respectively;  $P < .001$ ), had a higher number of comorbidities (0.91 [95% CI, 0.83-0.98] vs 0.80 [95% CI, 0.66-0.95];  $P = .008$ ), and made more frequent primary care practice visits (71.0% [95% CI, 68.5%-73.5%] vs 61.8% [95% CI, 57.3%-66.3%] with  $>3$  visits/year;  $P = .003$ ). Top tertile primary care physicians compared with the bottom tertile physicians had fewer minority patients (13.7% [95% CI, 10.6%-16.7%] vs 25.6% [95% CI, 20.2%-31.1%], respectively;  $P < .001$ ), non-English-speaking patients (3.2% [95% CI, 0.7%-5.6%] vs 10.2% [95% CI, 5.5%-14.9%];  $P < .001$ ), and patients with Medicaid coverage or without insurance (9.6% [95% CI, 7.5%-11.7%] vs 17.2% [95% CI, 13.5%-21.0%];  $P < .001$ ). After accounting for practice site and visit frequency differences, adjusting for patient panel factors resulted in a relative mean change in physician rankings of 7.6 percentiles (95% CI, 6.6-8.7 percentiles) per primary care physician, with more than one-third (36%) of primary care physicians (59/162) reclassified into different quality tertiles.

**Conclusion** Among primary care physicians practicing within the same large academic primary care system, patient panels with greater proportions of underinsured, minority, and non-English-speaking patients were associated with lower quality rankings for primary care physicians.

JAMA. 2010;304(10):1107-1113

www.jama.com

formance rankings that could have implications on how physicians are rewarded and on how resources are allocated within health care systems.

**Author Affiliations** are listed at the end of this article.  
**Corresponding Author:** Clemens S. Hong, MD, MPH, General Medicine Division, Massachusetts General Hospital, 50 Staniford St, Ninth Floor, Boston, MA 02115 (cshong@partners.org).

We tested the hypothesis that a physician's patient panel characteristics are independently associated with changes in his or her relative quality ranking. We ranked all primary care physicians within a single, academic primary care network according to a composite of commonly used Health Plan Employer and Data Information Set (HEDIS) measures and (1) compared patient panel characteristics of highest vs lowest ranked physicians, (2) examined changes in primary care physician rankings after adjusting for differences in patient panel characteristics, and (3) compared the patient panel characteristics of primary care physicians who moved up or down in their quality rankings.

## METHODS

We conducted a cohort analysis to study the relationship between patient panel characteristics and physician clinical performance. The Massachusetts General Hospital Practice-based research network includes 181 primary care physicians working in 9 hospital-affiliated practices and 4 community health centers.<sup>9</sup> All practices use the same electronic billing and scheduling systems and share an advanced electronic medical record system. Physicians are hired and credentialed using similar criteria, share the same compensation plan, and have similar staffing resources.

We identified 164 283 adult patients who visited any Massachusetts General Hospital Practice-based research network primary care practice from January 1, 2003, to December 31, 2005, using electronic billing records and excluding those who died (n=2817, which was determined based on review of social security records). We further excluded patients of physicians who had panels of less than 50 patients, those patients who were not linked to a primary care physician or Massachusetts General Hospital Practice-based research network primary care practice (n=36 546),<sup>9,10</sup> and those patients with missing data (n=2434). Our final analytic cohort consisted of

162 primary care physicians caring for 125 303 adult primary care patients.

We obtained data from an electronic record repository for Massachusetts General Hospital and affiliated institutions.<sup>11</sup> Patient characteristics included date of birth, sex, self-identified race/ethnicity, primary language, and insurance status. We assessed race/ethnicity data because prior studies have demonstrated associations with patient quality outcomes.<sup>12-19</sup> Based on a previously published method,<sup>20</sup> we estimated patient medical complexity using the number of chronic medical conditions for each patient taken from *International Classification of Diseases, Ninth Revision*, billing codes for 9 common chronic conditions: atrial fibrillation, chronic obstructive pulmonary disease, coronary artery disease, depression, diabetes, heart failure, hypertension, osteoarthritis, and stroke.

We also approximated socioeconomic status in the absence of patient-level administrative data,<sup>21</sup> by matching patient addresses to the geographic information system coordinates to obtain median household income and high school graduation rates for each patient's US Census block group. We matched addresses using StreetMap Premium software (Esri, Redlands, California), obtained geographic information system coordinates from successfully matched addresses using ArcGIS software (Esri), and input these geographic information system coordinates into the Demographic Update (Esri) to obtain 2007 US Census block group data. We matched 124 419 of the addresses (99.3%) with 80% sensitivity (requiring 80% of the characters to match identically) and an additional 838 using lower sensitivity limits (0.6%), leaving 46 unmatched (0.4%).

We created a composite quality measure based on 9 HEDIS measures including (1) mammography in the previous 2 years for eligible women aged 42 to 69 years; (2) Papanicolaou cervical screening in the previous 3 years for eligible women aged 21 to 64 years; (3) colonoscopy within 10 years, sigmoidoscopy or double-contrast barium

enema within 5 years, or home fecal occult blood testing within 1 year for eligible patients aged 52 to 69 years; (4) hemoglobin A<sub>1c</sub> testing in the prior year and proportion with levels of 7.0% or less in patients with diabetes; (5) low-density lipoprotein cholesterol testing in the previous year and proportion with levels of 100 mg/dL or less (to convert to mmol/L, multiply by 0.0259) for patients with diabetes and coronary artery disease.

To create the composite physician quality score, we first ranked primary care physicians for each of the 9 measures based on the log odds of achieving the performance measure using multilevel logistic regression models. This accounted for both clustering of patients within primary care physicians and for heteroscedasticity produced by the varying number of quality measurements for each primary care physician. We ranked primary care physicians after each of the 3 following stages of adjustment. Model 1 was unadjusted, model 2 was adjusted for practice and visit frequency, and model 3 was adjusted for practice, visit frequency, and all patient panel variables (patient age, sex, number of comorbidities, race/ethnicity, primary language spoken, and insurance status). In model 2, we first adjusted for primary care physician practice site to account for unmeasured practice characteristics and number of practice visits to allow for fairer assessment of primary care physician quality by adjusting for differences in direct patient contact (and therefore differences in direct opportunities to order the indicated tests or make appropriate management changes to achieve optimal hemoglobin A<sub>1c</sub> or low-density lipoprotein cholesterol control). Geo-coded variables for median household income and high school graduation rates were excluded in the final model because of minimal additional effect after adjustment for patient-level variables (mean change in primary care physician ranking of 1.9 percentiles).

We calculated physician composite rankings at each stage of adjustment as

the average of all individual scores for each of the 9 available HEDIS measures. The final composite score was converted to percentiles (range 1-100) with a lower percentile indicating higher primary care physician quality ranking. Each primary care physician thus had a composite quality score after each of the 3 stages of adjustment.

Using unadjusted composite quality scores (model 1), we compared physician and patient panel characteristics between physicians ranked in the top vs bottom quality tertile using the  $\chi^2$  test, the Wilcoxon rank sum test, or the *t* test. Then, we examined changes in relative physician rankings from the unadjusted base model (model 1) attributable to patient panel differences (model 3). All variables except for the intercept were considered as fixed effects in the multilevel regression analysis.

At each stage of adjustment, we identified physicians with a greater than 5 and 10 percentile absolute change in their relative composite quality ranking. Because many performance incen-

tive programs focus on tiers of physician quality, we also divided the 162 primary care physicians into tertiles based on their unadjusted composite quality rankings and assessed for reclassification into different postadjustment tertiles. We chose to classify physicians by tertiles because division into 3 quality categories is more intuitive than 2, and because it decreases the opportunity for misclassification, thus serving as a conservative classification approach. Finally, among the subset of physicians with a greater than 10 percentile change in their relative composite quality rankings from model 1 (unadjusted) to model 3 (fully adjusted), we used  $\chi^2$ , Wilcoxon rank sum, and *t* tests to compare the physician and patient panel characteristics of physicians who went up vs down in their composite quality rankings.

We used a threshold *P* value of less than .05 to determine statistical significance. Data were missing for less than 1.8% of all variables, and we performed a complete case analysis. We

used SAS statistical software version 9.1.3 (SAS Institute Inc, Cary, North Carolina) for all analyses except for multilevel models that were estimated using MLwiN multilevel modeling software version 2.11 (Centre for Multilevel Modelling, London, England). The Massachusetts General Hospital institutional review board approved the study.

## RESULTS

The 162 primary care physicians eligible for analysis were a mean of 18.6 years (95% confidence interval [CI], 16.9-20.2 years) from medical school graduation (TABLE 1). Fewer top tertile compared with bottom tertile primary care physicians practiced in community health centers (17.0% [95% CI, 6.9%-27.1%] vs 47.2% [95% CI, 33.7%-60.6%], respectively; *P* = .001). A greater proportion of female compared with male primary care physicians were in the top tertile of unadjusted primary care physician composite ranking (62.3% [95% CI, 49.2%-75.3%] vs 34.0% [95% CI, 21.2%-46.7%], respec-

**Table 1.** Primary Care Physician (PCP) and Patient Characteristics

	Mean (95% CI) <sup>a</sup>			P Value <sup>b</sup>
	PCPs Overall (N=162)	PCPs by Unadjusted Composite Quality Score Tertile		
		Top (n=53)	Bottom (n=53)	
<b>PCPs</b>				
Composite score percentile <sup>c</sup>	50.5 (46.0-55.0)	17.1 (14.4-19.7)	84.0 (81.4-86.6)	<.001
Period since medical school graduation, y	18.6 (16.9-20.2)	18.8 (16.3-21.3)	17.3 (14.2-20.5)	.17
Community health center practice <sup>d</sup>	48 (29.6) [22.6-36.7]	9 (17.0) [6.9-27.1]	25 (47.2) [33.7-60.6]	.001
Female sex <sup>d</sup>	79 (48.8) [41.1-56.5]	33 (62.3) [49.2-75.3]	18 (34.0) [21.2-46.7]	.004
Patient panel size	773 (706-841)	746 (631-861)	832 (700-965)	.40
<b>Patients (n = 125 303)</b>				
Panel age, y	49.2 (47.9-50.4)	51.1 (49.6-52.6)	46.6 (43.8-49.5)	<.001
No. of comorbidities	0.85 (0.79-0.91)	0.91 (0.83-0.98)	0.80 (0.66-0.95)	.008
Female sex, %	40.3 (36.6-44.0)	34.2 (27.6-40.8)	47.5 (42.0-53.0)	.002
Minority race/ethnicity, %	19.2 (16.7-21.7)	13.7 (10.6-16.7)	25.6 (20.2-31.1)	<.001
Non-English-speaking, %	6.3 (4.3-8.3)	3.2 (0.7-5.6)	10.2 (5.5-14.9)	<.001
Privately insured, %	69.7 (67.2-72.3)	72.4 (69.5-75.3)	66.0 (60.4-71.6)	.25
Medicare, %	17.5 (15.4-19.6)	18.0 (16.0-20.1)	16.7 (11.7-21.7)	.009
Medicaid or uninsured, %	12.8 (11.0-14.5)	9.6 (7.5-11.7)	17.2 (13.5-21.0)	<.001
High school graduate, % <sup>e</sup>	85.7 (84.5-86.9)	87.9 (86.5-89.4)	82.7 (80.3-85.1)	<.001
Family household income in US \$ <sup>e</sup>	60 155 (58 178-62 131)	63 901 (61 118-66 684)	53 890 (50 518-57 261)	<.001
>3 Primary care practice visits/y, %	67.3 (65.3-69.3)	71.0 (68.5-73.5)	61.8 (57.3-66.3)	.003

Abbreviation: CI, confidence interval.

<sup>a</sup>Unless otherwise indicated.

<sup>b</sup>Calculated using the  $\chi^2$  test, the Wilcoxon rank sum test, or the *t* test.

<sup>c</sup>Based on 9 Healthcare Effectiveness Data and Information Set measures. All PCPs were ranked and given percentile scores (a lower score indicates higher overall quality).

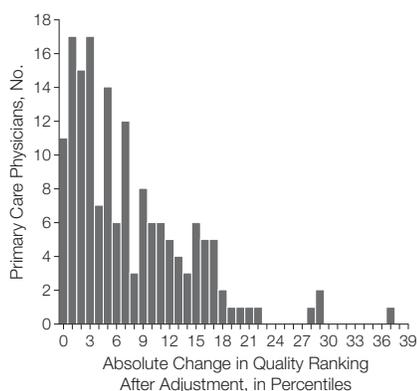
<sup>d</sup>Values are expressed as number (percentage) [95% CI].

<sup>e</sup>Based on 2007 US Census block group data.

**Table 2.** Changes in Primary Care Physician (PCP) Rankings on Composite Quality Scores

	Model 2 <sup>a</sup>	Model 3 <sup>b</sup>
Change in PCP rankings, mean percentile (95% CI)	15.1 (12.9-17.3)	7.6 (6.6-8.7)
PCPs with change in rankings, No. (%)		
>5 percentiles	122 (75.3)	95 (58.6)
>10 percentiles	86 (53.1)	52 (32.1)

Abbreviation: CI, confidence interval.

<sup>a</sup>Adjusted for practice and visit frequency.<sup>b</sup>Additionally adjusted for age, sex, comorbidity, race/ethnicity, language, and insurance status.**Figure.** Absolute Ranking Change for Primary Care Physicians From Practice- and Visit Frequency-Adjusted Model After Additional Adjustment for Patient Panel Characteristics

tively;  $P = .004$ ). The mean eligible patient panel size was 773 patients (95% CI, 706-841 patients).

Based on unadjusted composite quality rankings, patients of top tertile physicians compared with patients of bottom tertile physicians were older (51.1 years [95% CI, 49.6-52.6] vs 46.6 years [95% CI, 43.8-49.5 years], respectively;  $P < .001$ ), had a higher number of comorbidities (0.91 [95% CI, 0.83-0.98] vs 0.80 [95% CI, 0.66-0.95];  $P = .008$ ), made more frequent primary care practice visits (71.0% [95% CI, 68.5%-73.5%] vs 61.8% [95% CI, 57.3%-66.3%] with  $>3$  visits/year;  $P = .003$ ), and were less often female (34.2% [95% CI, 27.6%-40.8%] vs 47.5% [95% CI, 42.0-53.0];  $P = .002$ ) (Table 1).

The proportion of minority patients (13.7% [95% CI, 10.6%-16.7%] vs 25.6% [95% CI, 20.2%-31.1%];  $P < .001$ ), non-English-speaking patients (3.2% [95% CI, 0.7%-5.6%] vs 10.2% [95% CI, 5.5%-14.9%];  $P < .001$ ), and patients with

Medicaid coverage or without insurance (9.6% [95% CI, 7.5%-11.7%] vs 17.2% [95% CI, 13.5%-21.0%];  $P < .001$ ) was significantly lower in the top vs bottom tertile, respectively, of primary care physicians. Patients of top vs bottom tertile primary care physicians also lived in neighborhoods with higher median household incomes (\$63 901 [95% CI, \$61 118-\$66 684] vs \$53 890 [95% CI, \$50 518-\$57 261]), respectively;  $P < .001$ ) and higher high school graduation rates (87.9% [95% CI, 86.5%-89.4%] vs 82.7 [95% CI, 80.3-85.1];  $P < .001$ ).

Adjustment for practice and visit frequency (model 2) led to marked changes in relative physician quality rankings with more than 75% of primary care physicians changing by more than 5 percentiles and more than half of primary care physicians changing by more than 10 percentiles following adjustment (Table 2). Physician rankings changed by an average of 15.1 percentiles (95% CI, 12.9-17.3 percentiles) following adjustment for practice and visit frequency.

Additional adjustment for patient panel characteristics led to further re-ranking of primary care physicians, with 58.6% of primary care physicians changing more than 5 percentiles and 32.1% changing more than 10 percentiles. The mean change in physician rankings attributable to further adjustment for patient characteristics was 7.6 percentiles (95% CI 6.6-8.7 percentiles). The Figure shows the distribution of absolute changes in rank among primary care physicians following this adjustment.

When comparing the fully adjusted model (model 3) with the model adjusted for practice and visit frequency

(model 2), 11.3% of primary care physicians originally in the top composite score tertile decreased in ranking to the middle tertile, 14.3% of bottom tertile primary care physicians increased in ranking to the middle tertile, and 25.0% of middle tertile primary care physicians moved into the top or bottom tertile (Table 3).

The 34 primary care physicians whose relative quality rankings increased by more than 10 percentiles after adjustment were more likely to be physicians practicing at community health centers and have larger overall panel sizes. These physicians were also more likely to have panels with a higher proportion of minority, non-English-speaking, and younger patients with fewer comorbidities compared with the 44 primary care physicians whose relative quality rankings decreased after adjustment (Table 4).

## COMMENT

We examined relative physician quality rankings using an aggregate of commonly used HEDIS measures in a cohort of primary care physicians working within a single integrated health system. Primary care physicians in the top tertile of measured quality were more likely to care for older patients with greater comorbidity who made more frequent visits to see a primary care physician. This finding is consistent with prior studies.<sup>20</sup> Because older patients with more comorbidities are often seen more frequently, they may have stronger relationships with their physicians, and physicians caring for such patients may have more opportunities to complete process measures. Also, in concordance with literature in health care disparities,<sup>12-19</sup> top tertile primary care physicians were less likely to care for minority, non-English-speaking, Medicaid, and uninsured patients compared with bottom tertile primary care physicians.

After accounting for practice site and visit frequency differences, adjusting for patient panel factors resulted in an additional relative mean rank change of

7.6 percentiles (95% CI, 6.6-8.7 percentiles) in physician quality rankings. Our research expands on prior studies demonstrating the effect of patient case-mix on hospital-level quality.<sup>21,22</sup> As seen in these studies, the magnitude of the change in physician rankings in our study was moderate for most primary care physicians and large for few; however, even modest changes in rankings have important consequences for physician reclassification between fixed thresholds (such as those often used in performance incentive programs and quality reporting). These changes in ranking after adjustment resulted in the reclassification of 36% of primary care physicians (59/162) into different tertiles of composite quality score.

Moreover, primary care physicians whose ranking increased by greater than 10 percentiles after adjustment were more likely to work in community health centers and care for a higher proportion of minority and non-English-speaking patients. Therefore, one potential risk of not adjusting for patient panel makeup is undervalued quality ranking for primary care physicians who work in community health centers or those who take care of minority and non-English-speaking patients. This illustrates one potential unintended consequence of current (ie, not adjusted for patient panel) physician clinical performance assessment when tied to performance-based incentives.

Our findings provide evidence of the effect of patient panel makeup on attainment of HEDIS quality measures and support our hypothesis that patient panel characteristics are associated with changes in relative physician quality scores. Furthermore, our study demonstrates possible effects of this finding on misclassification of physicians when ranking or tiers are used to compare physician quality and reveals a potential mechanism for unintended consequences posed by physician clinical performance assessment in the setting of performance incentive programs.

These results have considerable policy implications given the increasing role of physician performance incentive programs.<sup>6</sup> Prior studies have posited numerous potential unintended consequences of these programs.<sup>22-27</sup> Our study demonstrates a potential mechanism through which performance incentive programs might worsen health care disparities. Because physicians and practices with higher quality scores receive higher payments and recognition under these programs, these incentive approaches could erroneously distribute resources away

from high-quality physicians caring for more vulnerable patients and worsen health care disparities.<sup>6,25,26</sup>

Additionally, prior studies by our group demonstrate that socioeconomically deprived patients are less likely to be connected to their physicians and make fewer visits to their primary care physicians.<sup>9</sup> These practice-connected rather than primary care physician-connected patients were more likely to be found in community health centers.<sup>9</sup> Beyond issues surrounding patient attribution, decreased connectedness to physicians could exacerbate the

**Table 3.** Changes in Primary Care Physician (PCP) Composite Score Tertile

Initial Tertile <sup>a</sup>	No. (%) of PCPs by Postadjustment Tertile <sup>b</sup>		
	1	2	3
1 (n=53)	47 (78.7)	6 (11.3)	0
2 (n=56)	6 (10.7)	42 (75.0)	8 (14.3)
3 (n=53)	0	8 (14.3)	45 (84.9)

<sup>a</sup>Adjusted for practice and visit frequency (model 2).

<sup>b</sup>Additionally adjusted for patient age, sex, comorbidity, race/ethnicity, language, and insurance status (model 3).

**Table 4.** Primary Care Physician (PCP) and Patient Characteristics From Unadjusted Baseline Model (Model 1) to Fully Adjusted Model (Model 3)

	Percentile Change in PCP Ranking, Mean (95% CI) <sup>a</sup>		P Value <sup>b</sup>
	>10 Increase (n=34)	>10 Decrease (n=44)	
<b>Physicians</b>			
Composite score percentile <sup>c</sup>	64.3 (55.8-72.8)	43.6 (37.3-49.9)	<.001
Period since medical school graduation, y	16.1 (12.5-19.7)	20.7 (17.0-24.3)	.08
Community health center practice <sup>d</sup>	16 (47.1) [30.3-63.8]	9 (20.5) [8.5-32.4]	.01
Female sex <sup>d</sup>	20 (58.8) [42.3-75.4]	15 (34.1) [20.1-48.1]	.03
Patient panel size	946 (804-1088)	740 (609-870)	.02
<b>Patients</b>			
Panel age, y	46.4 (42.9-49.9)	51.3 (49.4-53.1)	<.001
No. of comorbidities	0.78 (0.61-0.96)	0.93 (0.82-1.03)	.01
Female sex, %	36.6 (29.1-44.1)	46.8 (38.8-54.7)	.08
Minority race/ethnicity, %	27.3 (19.4-35.1)	14.4 (11.2-17.7)	.002
Non-English-speaking, %	11.1 (4.8-17.4)	4.4 (1.8-7.0)	.003
Privately insured, %	67.2 (59.8-74.6)	70.6 (66.6-74.5)	.80
Medicare, %	16.2 (9.8-22.5)	18.8 (15.7-21.8)	.02
Medicaid or uninsured, %	16.7 (11.8-21.5)	10.7 (8.0-13.4)	.06
High school graduate, % <sup>e</sup>	17.5 (14.1-20.8)	12.7 (10.9-14.5)	.006
Family household income in US \$ <sup>e</sup>	54 618 (49 999-59 237)	63 241 (59 762-66 719)	<.001
>3 Primary care practice visits/y, %	66.3 (61.1-71.4)	69.4 (66.9-72.0)	.90

Abbreviation: CI, confidence interval.

<sup>a</sup>Unless otherwise indicated.

<sup>b</sup>Calculated using the  $\chi^2$  test, the Wilcoxon rank sum test, or the *t* test.

<sup>c</sup>Based on 9 Healthcare Effectiveness Data and Information Set measures. All PCPs were ranked and given percentile scores (a lower score indicates higher overall quality).

<sup>d</sup>Values are expressed as number (percentage) [95% CI].

<sup>e</sup>Based on 2007 US Census block group data.

effect of inadequate adjustment for patient characteristics, leading to even lower incentive-based reimbursement for physicians working in community health centers and other safety-net sites. This reduction in income could lead to a reduction in both the number of physicians working in such communities and practice-level resources available to invest in structures and processes to improve quality. Physicians in these areas would be at an additional disadvantage due to traditionally decreased access to resources and a payer mix that is likely to include a higher proportion of uninsured patients and Medicaid recipients. Further studies using real-world performance incentive models may help us understand the effect on redistribution of resources within and between health care systems.

Our aim was not to develop a method for case-mix adjustment, but to study the effect of patient characteristics on physician clinical performance ranking. Further work in designing appropriate case-mix adjustment methods for physician clinical performance assessment is necessary. Case-mix adjustment may help improve physician acceptance of physician-level profiles by leveling the playing field and attenuating the effect of potential unintended consequences, and decrease the chance that physicians will seek to exclude patients likely to worsen their measured quality. These adjustment methods should consider the effect of the practice characteristics, including resources and infrastructure.<sup>28,29</sup> However, because visit frequency is at least partially due to physician practice style, it should be excluded and further work is warranted to identify best practice follow-up appointment intervals for different clinical scenarios. Lastly, novel methods and performance incentive schemes must be paired with ongoing stratification of quality performance by sociodemographic characteristics to avoid giving undue credit to physicians and practices for lower performance with vulnerable patients and ensure that disparities in care are made the target of performance incentive pro-

grams rather than obscured by case-mix adjustment practices.<sup>14,17</sup>

Our study strengths included the use of easily obtained patient characteristics and widely used clinical quality measures (drawn from system databases) that represent actual receipt of care rather than reported data. Furthermore, comparing physicians using an aggregate quality measure represents an overall quality construct, allows a greater number of patients to be included, and minimizes the influence of any single measure-specific outcome.<sup>30</sup> The compositing method used in our study allows for easy addition of measures as well as transparent weighting when necessary. However, despite the advantages of such composite quality measures, few quality measurement tools incorporate other dimensions of clinical quality (eg, physician empathy and communication skills), and it remains important to continue evaluating physician clinical performance on individual measures because quality improvement is more easily targeted at specific outcomes and because one measure may not predict performance on others.<sup>31-33</sup>

We must interpret our results within the context of the study design. While it is a strength to compare the association of patient characteristics on clinical performance measures among a relatively homogenous physician cohort, our results may not be generalizable to other health care systems such as community health systems or smaller private practice networks that makeup the majority of health care delivery in the United States.<sup>34</sup> In addition, use of currently available quality measures that are not comprehensive, are process-oriented, and may not measure actual physician quality has limitations. Nonetheless, because these HEDIS-based measures are widely used, our findings remain directly applicable to current quality measurement practices. Another consideration is that quality estimates for some physicians may be more reliable for one measure than another.<sup>35,36</sup> Although we addressed

heterogeneity in sample sizes between physicians using multilevel modeling techniques, this continues to be a central hurdle for physician-level clinical performance measures. Finally, although adjustment for patient variables in this study may have contributed to a more accurate estimation of quality outcomes, they only account for a fraction of the total variability in physician-level quality scores. Additional work is necessary to develop a wider range of reliable patient-, physician-, and practice-level variables.

In summary, our study demonstrates that patient panel characteristics are associated with the relative measured quality of physicians within a large academic primary care network. Adjustment for differences in patient panel characteristics resulted in significant reclassification of top tier vs bottom tier physicians. To the extent that health systems reward physicians for higher measured quality of care, lack of adjustment for patient panel characteristics may penalize physicians for taking care of more vulnerable patients, incentivize physicians to select patients to improve their quality scores, and result in the misallocation of resources away from physicians taking care of more vulnerable populations. Conversely, adjustment for patient panel characteristics may remove the incentive to improve care or may inappropriately reward lower-quality physicians caring for more vulnerable patients. Efforts to improve quality of care must address both fairness of physician clinical performance assessment and the design of incentive schemes to both provide equitable distribution of resources and reduce disparities in care for vulnerable patients.

**Author Affiliations:** Schools of Medicine (Drs Hong, Atlas, Chang, Barry, and Grant) and Public Health (Dr Subramanian), Harvard University, Boston, Massachusetts; and General Medicine Division, Massachusetts General Hospital, Boston (Drs Hong, Atlas, Chang, Barry, and Grant, and Mr Ashburner).

**Author Contributions:** Dr Hong had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Study concept and design:** Hong, Atlas, Chang, Subramanian, Barry, Grant.

**Acquisition of data:** Hong, Atlas, Ashburner.

**Analysis and interpretation of data:** Hong, Atlas, Chang, Subramanian, Ashburner, Grant.  
**Drafting of the manuscript:** Hong, Grant.  
**Critical revision of the manuscript for important intellectual content:** Hong, Atlas, Chang, Subramanian, Ashburner, Barry, Grant.  
**Statistical analysis:** Hong, Chang, Subramanian, Ashburner.  
**Obtained funding:** Grant.  
**Administrative, technical, or material support:** Hong, Atlas, Ashburner.  
**Study supervision:** Atlas, Subramanian, Grant.  
**Financial Disclosures:** None reported.

**Funding/Support:** Dr Hong is supported by National Research Service Award grant T32 HP12706 from the National Institutes of Health. Dr Grant is supported by career development grant NIDDK K23 DK067452 from the National Institutes of Health. Dr Atlas is supported by grant R18 HS018161 from the Agency for Healthcare Research and Quality. Dr Subramanian is supported by career development grant NHLBI K25 HL081275 from the National Institutes of Health.  
**Role of the Sponsor:** The National Institutes of Health did not participate in the design and conduct of the study; in the collection, analysis, and interpretation of

the data; or in the preparation, review, or approval of the manuscript.

**Additional Contributions:** Timothy Ferris, MD, MPH, the medical director of the Massachusetts General Physicians Organization and associate professor of medicine at Harvard Medical School, provided helpful discussions on the manuscript. Dr Ferris did not receive any compensation for his contribution. Guoping Huang from the Harvard University Center for Geographic Analysis helped us create the US Census block group data on socioeconomic status using geo-codes. The Harvard University Center for Geographic Analysis was paid a small fee for these services.

## REFERENCES

- 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(26):2635-2645.
- Fisher ES, Wennberg DE, Stukel TA, et al. The implications of regional variations in Medicare spending, part 2: health outcomes and satisfaction with care. *Ann Intern Med*. 2003;138(4):288-298.
- Fisher ES, Wennberg DE, Stukel TA, et al. The implications of regional variations in Medicare spending, part 1: the content, quality, and accessibility of care. *Ann Intern Med*. 2003;138(4):273-287.
- Rosenthal MB, Landon BE, Normand SL, et al. Pay for performance in commercial HMOs. *N Engl J Med*. 2006;355(18):1895-1902.
- Rosenthal MB, Fernandez R, Song HR, Landon B. Paying for quality: providers' incentives for quality improvement. *Health Aff (Millwood)*. 2004;23(2):127-141.
- Chien AT, Chin MH, Davis AM, Casalino LP. Pay for performance, public reporting, and racial disparities in health care: how are programs being designed? *Med Care Res Rev*. 2007;64(5)(suppl):283S-304S.
- Rosenthal MB, Frank RG, Li Z, Epstein AM. Early experience with pay-for-performance: from concept to practice. *JAMA*. 2005;294(14):1788-1793.
- Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. *JAMA*. 2005;293(10):1239-1244.
- Atlas SJ, Grant RW, Ferris TG, Chang Y, Barry MJ. Patient-physician connectedness and quality of primary care. *Ann Intern Med*. 2009;150(5):325-335.
- Atlas SJ, Chang Y, Lasko TA, et al. Is this "my" patient? development and validation of a predictive model to link patients to primary care providers. *J Gen Intern Med*. 2006;21(9):973-978.
- Murphy SN, Chueh HC. A security architecture for query tools used to access large biomedical databases. *Proc AMIA Symp*. 2002;552-556.
- Smedley BD, Stith AY, Nelson AR. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: Institute of Medicine; 2003.
- Virnig BA, Lurie N, Huang Z, et al. Racial variation in quality of care among Medicare+Choice enrollees. *Health Aff (Millwood)*. 2002;21(6):224-230.
- Schneider EC, Zaslavsky AM, Epstein AM. Racial disparities in the quality of care for enrollees in medicare managed care. *JAMA*. 2002;287(10):1288-1294.
- Schneider EC, Cleary PD, Zaslavsky AM, Epstein AM. Racial disparity in influenza vaccination: does managed care narrow the gap between African Americans and whites? *JAMA*. 2001;286(12):1455-1460.
- McBean AM, Huang Z, Virnig BA, Lurie N, Musgrave D. Racial variation in the control of diabetes among elderly medicare managed care beneficiaries. *Diabetes Care*. 2003;26(12):3250-3256.
- Fiscella K, Franks P, Gold MR, Clancy CM. Inequality in quality: addressing socioeconomic, racial, and ethnic disparities in health care. *JAMA*. 2000;283(19):2579-2584.
- Martin LM, Calle EE, Wingo PA, Heath CW Jr. Comparison of mammography and Pap test use from the 1987 and 1992 National Health Interview Surveys: are we closing the gaps? *Am J Prev Med*. 1996;12(2):82-90.
- Pearlman DN, Rakowski W, Ehrich B, Clark MA. Breast cancer screening practices among black, Hispanic, and white women: reassessing differences. *Am J Prev Med*. 1996;12(5):327-337.
- Higashi T, Wenger NS, Adams JL, et al. Relationship between number of medical conditions and quality of care. *N Engl J Med*. 2007;356(24):2496-2504.
- Zaslavsky AM, Epstein AM. How patients' socio-demographic characteristics affect comparisons of competing health plans in California on HEDIS quality measures. *Int J Qual Health Care*. 2005;17(1):67-74.
- Mehta RH, Liang L, Karve AM, et al. Association of patient case-mix adjustment, hospital process performance rankings, and eligibility for financial incentives. *JAMA*. 2008;300(16):1897-1903.
- Fisher ES. Paying for performance—risks and recommendations. *N Engl J Med*. 2006;355(18):1845-1847.
- Karve AM, Ou FS, Lytle BL, Peterson ED. Potential unintended financial consequences of pay-for-performance on the quality of care for minority patients. *Am Heart J*. 2008;155(3):571-576.
- Casalino LP, Elster A, Eisenberg A, et al. Will pay-for-performance and quality reporting affect health care disparities? *Health Aff (Millwood)*. 2007;26(3):w405-w414.
- 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. *J Gen Intern Med*. 2009;24(7):854-859.
- Salem-Schatz S, Moore G, Rucker M, Pearson SD. The case for case-mix adjustment in practice profiling: when good apples look bad. *JAMA*. 1994;272(11):871-874.
- Friedberg MW, Coltin KL, Safran DG, et al. Associations between structural capabilities of primary care practices and performance on selected quality measures. *Ann Intern Med*. 2009;151(7):456-463.
- Poon EG, Wright A, Simon SR, et al. Relationship between use of electronic health record features and health care quality: results of a statewide survey. *Med Care*. 2010;48(3):203-209.
- Reeves D, Campbell SM, Adams J, et al. Combining multiple indicators of clinical quality: an evaluation of different analytic approaches. *Med Care*. 2007;45(6):489-496.
- Gandhi TK, Francis EC, Puopolo AL, et al. Inconsistent report cards: assessing the comparability of various measures of the quality of ambulatory care. *Med Care*. 2002;40(2):155-165.
- Rosenthal GE. Weak associations between hospital mortality rates for individual diagnoses: implications for profiling hospital quality. *Am J Public Health*. 1997;87(3):429-433.
- Wilson IB, Landon BE, Marsden PV, et al. Correlations among measures of quality in HIV care in the United States: cross sectional study. *BMJ*. 2007;335(7629):1085.
- Gonzalez ML. *Socioeconomic Characteristics of Medical Practice 1997*. Chicago, IL: American Medical Association; 1997.
- Greenfield S, Kaplan SH, Kahn R, Ninomiya J, Griffith JL. Profiling care provided by different groups of physicians: effects of patient case-mix (bias) and physician-level clustering on quality assessment results. *Ann Intern Med*. 2002;136(2):111-121.
- Kaplan SH, Griffith JL, Price LL, Pawlson LG, Greenfield S. Improving the reliability of physician performance assessment: identifying the "physician effect" on quality and creating composite measures. *Med Care*. 2009;47(4):378-387.