Annals of Internal Medicine ORIGINAL RESEARCH

The Value-Based Payment Modifier: Program Outcomes and Implications for Disparities

Eric T. Roberts, PhD; Alan M. Zaslavsky, PhD; and J. Michael McWilliams, MD, PhD

Background: When risk adjustment is inadequate and incentives are weak, pay-for-performance programs, such as the Value-Based Payment Modifier (Value Modifier [VM]) implemented by the Centers for Medicare & Medicaid Services, may contribute to health care disparities without improving performance on average.

Objective: To estimate the association between VM exposure and performance on quality and spending measures and to assess the effects of adjusting for additional patient characteristics on performance differences between practices serving higher-risk and those serving lower-risk patients.

Design: Exploiting the phase-in of the VM on the basis of practice size, regression discontinuity analysis and 2014 Medicare claims were used to estimate differences in practice performance associated with exposure of practices with 100 or more clinicians to full VM incentives (bonuses and penalties) and exposure of practices with 10 or more clinicians to partial incentives (bonuses only). Analyses were repeated with 2015 claims to estimate performance differences associated with a second year of exposure above the threshold of 100 or more clinicians. Performance differences were assessed between practices serving higher- and those serving lower-risk patients after standard Medicare adjustments versus adjustment for additional patient characteristics.

Setting: Fee-for-service Medicare.

Patients: Random 20% sample of beneficiaries.

Measurements: Hospitalization for ambulatory care–sensitive conditions, all-cause 30-day readmissions, Medicare spending, and mortality.

Results: No statistically significant discontinuities were found at the threshold of 10 or more or 100 or more clinicians in the relationship between practice size and performance on quality or spending measures in either year. Adjustment for additional patient characteristics narrowed performance differences by 9.2% to 67.9% between practices in the highest and those in the lowest quartile of Medicaid patients and Hierarchical Condition Category scores.

Limitation: Observational design and administrative data.

Conclusion: The VM was not associated with differences in performance on program measures. Performance differences between practices serving higher- and those serving lower-risk patients were affected considerably by additional adjustments, suggesting a potential for Medicare's pay-for-performance programs to exacerbate health care disparities.

Primary Funding Source: The Laura and John Arnold Foundation and National Institute on Aging.

Ann Intern Med. 2018;168:255-265. doi:10.7326/M17-1740 **Annals.org**
For author affiliations, see end of text.
This article was published at Annals.org on 28 November 2017.

In January 2017, the Centers for Medicare & Medicaid Services (CMS) implemented the Merit-based Incentive Payment System (MIPS), establishing a new payment system for clinicians in the fee-for-service Medicare program (1). As part of a broader push to link provider payments to value (2–5), the MIPS is a pay-for-performance program that intends to reward clinicians for improving quality of care and reducing spending.

Although the effects of this program will not be known for several years, its basic design is similar to that of its predecessor, the Value-Based Payment Modifier (Value Modifier [VM]) (6). Each year from 2013 through 2016, the VM assessed the performance of physician practices on a set of quality and spending measures and adjusted Part B payment rates in the Medicare Physician Fee Schedule 2 years later on the basis of these performance scores (7).

In 2013, practices with 100 or more clinicians were required to meet reporting requirements or incur a small reduction in 2015 payment rates, but exposure to the VM (that is, performance-based payment adjustments) was optional (8). In 2014, the VM became mandatory for all practices with 10 or more clinicians, except those participating in alternative payment models, such as Medicare's Accountable Care Organization (ACO) programs (9). Practices with 100 or more clinicians were subject to upward, downward, or neutral performance-based payment adjustments, those with 10 to 99 were subject to upward or neutral—but not downward—adjustments, and those with fewer than 10 were unaffected (3, 10). In 2015, all practices with 10 or more clinicians were exposed to full VM incentives (both penalties and bonuses) (11). Base payment adjustments ranged from −2% to 2% on the basis of 2014 performance and from −4% to 4% on the basis of 2015 performance, but high-performing practices have received much higher bonuses (for example, rate increases of 16% to 32% in 2016), because the VM's budget neutrality provision stipulated that penalties for failing to meet reporting requirements be redistributed as bonuses (12, 13).

See also:

Editorial comment 291

Web-Only
Supplement

To date, many performance measures used in the VM and MIPS have been adjusted for only a limited set of patient characteristics (14–16), raising concerns that practices' performance scores may partly reflect differences in their patients' clinical or social characteristics, rather than only differences in quality of care (17–22). Because budget neutrality provisions in these programs require penalties and bonuses to offset, inadequate risk adjustment might result in sustained and unwarranted transfers of resources from practices serving sicker or more socially disadvantaged patients to those serving healthier or more affluent patients (23–27).

In evaluating the merits of pay-for-performance programs, it therefore is important to consider both the behavioral response elicited by program incentives and the implications of inadequate risk adjustment for health care disparities. In this study, we assessed differences in performance on quality and spending measures associated with the exposure of practices with 10 or more and those with 100 or more clinicians to partial or full VM incentives in 2014, respectively, as well as performance differences associated with the exposure of practices with 100 or more clinicians to a second year of incentives in 2015. In a second set of analyses, we examined the effect of adjusting for additional patient characteristics on practice rankings and on performance differences between practices with larger proportions of low-income and medically complex patients and practices with smaller proportions of such patients.

METHODS

Study Design

For our first set of analyses, we used a cross-sectional regression discontinuity design to assess differences in spending and quality between practices above and below the size thresholds determining exposure to the VM. This design exploits the fact that exposure to performance incentives in the VM differed between practices above and below specific thresholds but other determinants of spending and quality likely did not, enabling an inference similar to that from a randomized study (28). Because too few observations may exist within a narrow range of a threshold to support comparisons, regression discontinuity studies typically use broader ranges of data and regression analysis to estimate discontinuities (that is, level shifts) in outcomes above versus below a threshold. Thus, we analyzed data from practices with 50 to 150 clinicians (for the threshold of 10 or more clinicians) and 2 to 30 clinicians (for thresholds of 10 or more clinicians) to estimate discontinuities in spending and quality. We assumed that the relationship between practice size and performance would have an uninterrupted approximately linear trend across these thresholds in the absence of the VM.

In our second set of analyses, we assessed practice performance before versus after adjusting for additional patient characteristics not included in the risk-adjustment methods used by CMS in the VM or MIPS (14, 15). We assessed performance differences between practices serving larger proportions of low-income and medically complex patients and those serving smaller proportions of such patients and compared these differences before and after the additional adjustments. We also assessed changes in the relative performance ranking of practices after the additional adjustments. This second set of analyses illustrates the implications of limited risk adjustment for health care disparities, not only in the VM but also in the MIPS.

Data Sources and Study Population

We analyzed claims and enrollment data in 2014 and 2015 for a random 20% sample of beneficiaries who were continuously enrolled in Parts A and B of fee-for-service Medicare in the year of interest (while alive in the case of decedents) and the preceding year (to assess established diagnoses). Following methods used by CMS for the VM, we attributed each beneficiary to the practice (defined by CMS as a taxpayer identification number [TIN]) that accounted for the largest share of allowed charges for that beneficiary's office visits during the study year (**Supplement**, available at Annals.org) (29). Beneficiaries without an office visit during the year (13%) were excluded. To exclude practices unaffected by the VM, we used CMS data on ACO participants to remove practices participating in the Pioneer model or Medicare Shared Savings Program (**Supplement**) (30).

Practice Exposure to the VM

To determine practice size and thus exposure to VM incentives, we used the 2014 Medicare Provider Practice and Specialty file to attribute each clinician to the TINs under which they billed for Part B services (**Supplement**) (31). We calculated the total number of clinicians billing under each TIN and created indicators for 3 size categories: fewer than 10 clinicians (no exposure), 10 to 99 clinicians (exposed to potential bonuses only), and 100 or more clinicians (exposed to potential bonuses and penalties). Each category had a different exposure to VM incentives in 2014. Our method for determining practice size closely followed the approach used by CMS to determine practice size for the VM and yielded a total number of practices with 100 or more clinicians that was very similar to that reported by CMS (**Supplement Table 1**, available at Annals.org) (6, 9, 32).

Outcome Variables

We examined 3 annual measures of quality and spending that CMS assessed as core performance measures for all practices subject to the VM: admissions for ambulatory care–sensitive conditions (ACSCs) (14), total Medicare Part A and Part B spending per beneficiary (15), and all-cause readmissions within 30 days of hospital discharge (**Supplement**) (16). Although annual mortality was not included as a performance measure in the VM, we assessed it as an additional measure that may be particularly sensitive to risk adjustment and can be interpreted as a health outcome more reliably than utilization-based quality measures (for example, admissions and readmissions often may be appropriate and improve health).

Patient Characteristics

We used Medicare enrollment data to determine age, sex, and race/ethnicity of beneficiaries; whether end-stage renal disease was present; whether beneficiaries were enrolled in Medicaid (dual eligibility); and whether disability was the original reason for Medicare entitlement. We used the Chronic Conditions Data Warehouse to determine the presence of 27 chronic conditions before each study year. Finally, for each beneficiary, we calculated a Hierarchical Condition Category (HCC) risk score based on enrollment information (including Medicaid coverage) and clinical diagnoses from claims in the preceding year (33).

Statistical Analysis

For each outcome, we conducted a regression discontinuity analysis to isolate differences associated with exposure of practices in 2014 to bonuses *and* penalties above the threshold of 100 or more clinicians (vs. only bonuses below the threshold) and to bonuses above the threshold of 10 or more clinicians (vs. neither bonuses nor penalties below the threshold). Specifically, we fitted a patient-level linear regression model to estimate the difference in performance between practices above each threshold and those below it, adjusting for the linear relationship between the outcome and practice size (number of clinicians) and for patients' clinical and sociodemographic characteristics (Supplement). This adjusted difference (or adjusted discontinuity) may be interpreted as the difference in performance attributable to VM incentives. We repeated our analysis of the threshold of 100 or more clinicians with data from 2015, when practices with 10 to 99 clinicians were also exposed to penalties, to isolate performance differences associated with 2 years of exposure to full VM incentives (above the threshold) versus 1 year of exposure to full incentives (below the threshold).

We conducted several analyses to test the assumptions of our regression discontinuity approach and to explore potential sources of bias. First, we tested for threshold-related discontinuities in the relationship between practice size and observable patient characteristics to determine whether thresholds were systematically associated with differences in patient populations. Second, we checked for evidence of bunching around thresholds in the distribution of practice size, which might suggest manipulation of size by practices to escape or gain exposure to the VM (Supplement) (34).

Third, we narrowed the range of practice sizes included in our analyses to focus on practices closer to the thresholds, thereby relaxing our assumption that the relationship between practice size and performance was linear (but at the expense of precision). Fourth, we conducted falsification tests by using arbitrary practice size thresholds unrelated to VM exposure and by repeating our analyses with data from 2012 (before VM implementation). Fifth, we excluded practices with 95 to 105 clinicians to minimize attenuation bias from minor inaccuracies in our measurement of practice size and from changes in practice size from 2014 to 2015. After this exclusion, only 5% of practices with 50 to 150 clinicians in 2014 moved above or below the threshold of 100 or more clinicians in 2015.

For our second set of analyses, we categorized practices into quartiles based on the proportion of beneficiaries in each practice who were dually enrolled in Medicaid (an indicator of qualifying disabilities or low income) and separately on the basis of the mean HCC score of patients in each practice. We estimated differences in performance between quartiles of practices after first adjusting for a base set of variables used by CMS for risk adjustment in the VM and MIPS and then after adjusting for additional patient characteristics we could assess from Medicare administrative data (Table 1).

To implement adjustments, we first used linear regression to estimate associations between patient characteristics and outcomes within practices, pooling data across practices. From these within-practice analyses, we predicted each practice's expected performance on the basis of the characteristics of its patients, ignoring the practice's distinct contribution to quality or spending (Supplement). We then adjusted estimates for each practice by subtracting expected performance from observed performance. If, for example, high-risk patients had worse outcomes than low-risk patients within practices, the effect of this within-practice difference would be removed by the adjustment. On the other hand, if high-risk patients disproportionately sorted to low-quality practices, this association would persist in the adjusted estimates.

In addition, we performed a simulation to assess the extent to which additional risk adjustment would affect practice rankings under pay-for-performance programs like the MIPS, which use a continuous scoring system to rank practices on the basis of their performance relative to other practices reporting on the same measures and determines bonuses and penalties from the rankings (Supplement) (35). On the basis of these changes in rankings, we also determined the proportion of practices expected to gain or lose eligibility for the MIPS exceptional-performance bonus for each measure (available to practices above the 62.5th percentile [36]) and the percentile changes expected for 5% of the most affected practices. Finally, we estimated the proportion that would have had changes in eligibility for VM bonuses or penalties (those performing ≥1 SD above or below the mean) from the additional adjustments (Supplement).

Because many performance measures in the VM and MIPS are not derived from claims data, we could not analyze changes in practice rankings and payment adjustments on the basis of the full composite quality scores calculated by the programs—we could do so only for each claims-based measure we analyzed. Assessments for individual measures are nevertheless instructive, because the expected effect of the additional adjustments on overall scores would be an average of the effects on constituent measures. To support computational efficiency and minimize measurement error

Table 1. Patient Characteristics Used to Risk-Adjust Practice Performance

Characteristic	Outcome		
	Hospitalizations for ACSCs*	**Total Annual Medicare Spending per Beneficiary***	**Mortality**
Base risk adjustment†	Age-by-sex categories	Age Sex HCC score‡ and HCC score squared End-stage renal disease	Age Sex HCC score‡ and HCC score squared End-stage renal disease
Additional patient characteristics	70 HCC indicators§ CCW conditions‖ End-stage renal disease Disability as original reason for Medicare enrollment Dual Medicare and Medicaid enrollment¶ Recipients of Medicare Savings Program** Interactions among variables††	CCW conditions‖ Recipients of Medicare Savings Program** Interactions among variables††	CCW conditions‖ Recipients of Medicare Savings Program** Interactions among variables††

ACSC = ambulatory care–sensitive condition; CCW = Chronic Conditions Data Warehouse; CMS = Centers for Medicare & Medicaid Services; HCC = Hierarchical Condition Category; VM = Value-Based Payment Modifier.
* Performance measure in the VM.
† For ACSC hospitalizations and spending, the base model included all patient-level variables used by CMS to risk-adjust these outcomes for the VM, in addition to state fixed effects. We used the variables of CMS adjustments for spending as the base adjustment variable set for mortality.
‡ Includes indicators of dual enrollment in Medicare and Medicaid and disability status.
§ Included in the CMS HCC risk-adjustment model. These are used to construct patients' HCC scores.
‖ Indicators for the presence of 27 chronic conditions reported before the study year, plus counts of chronic conditions (indicator variables in unit increments between 2 and 8 conditions and for ≥9 conditions).
¶ Includes low-income persons aged <65 y who qualified for Medicare because of a disability and persons aged ≥65 y who additionally qualified for Medicaid because of low income.
** Includes partial Medicaid enrollees in the Qualified Medicare Beneficiary, Specified Low-Income Medicare Beneficiary, and Qualifying Individual programs.
†† Two-way interactions between HCC score, count of prior-year CCW chronic conditions, disability status, dual Medicare and Medicaid enrollment, and recipients of the Medicare Savings Program. To account for differences in eligibility for full Medicaid coverage across states, we included interactions between state fixed effects and a patient-level indicator of full dual Medicare and Medicaid enrollment.

in classifying practices into quartiles based on dual eligibility and HCC scores, we focused our second set of analyses on practices with 100 or more clinicians and excluded readmissions as an outcome, because this measure would have further restricted the sample to hospitalized patients (22% of the sample).

Role of the Funding Source

The funding sources had no role in the design, conduct, or reporting of the study.

RESULTS

Regression Discontinuity Analysis

After adjustment for patient characteristics and the linear relationship with practice size, differences in hospitalization for ACSCs, readmissions, Medicare spending, and mortality between practices above the size thresholds and those below (the adjusted discontinuities) were small in 2014 and not statistically significant (Figure 1). For example, exceeding the threshold of 10 or more clinicians was associated with an average of 0.0027 more hospitalizations for ACSCs per beneficiary (95% CI, −0.0003 to 0.0056 [P = 0.078]), and exceeding the threshold of 100 or more clinicians was associated with an average of 0.002 fewer ACSC hospitalizations per beneficiary (CI, 0.006 to 0.003 fewer hospitalizations [P = 0.48]) than expected from the linear relationship between admission rates and practice size. Analyses of the threshold of 100 or more clinicians that used 2015 data revealed no statistically significant discontinuities associated with a second year of full exposure to the VM (Supplement Table 2, available at Annals.org).

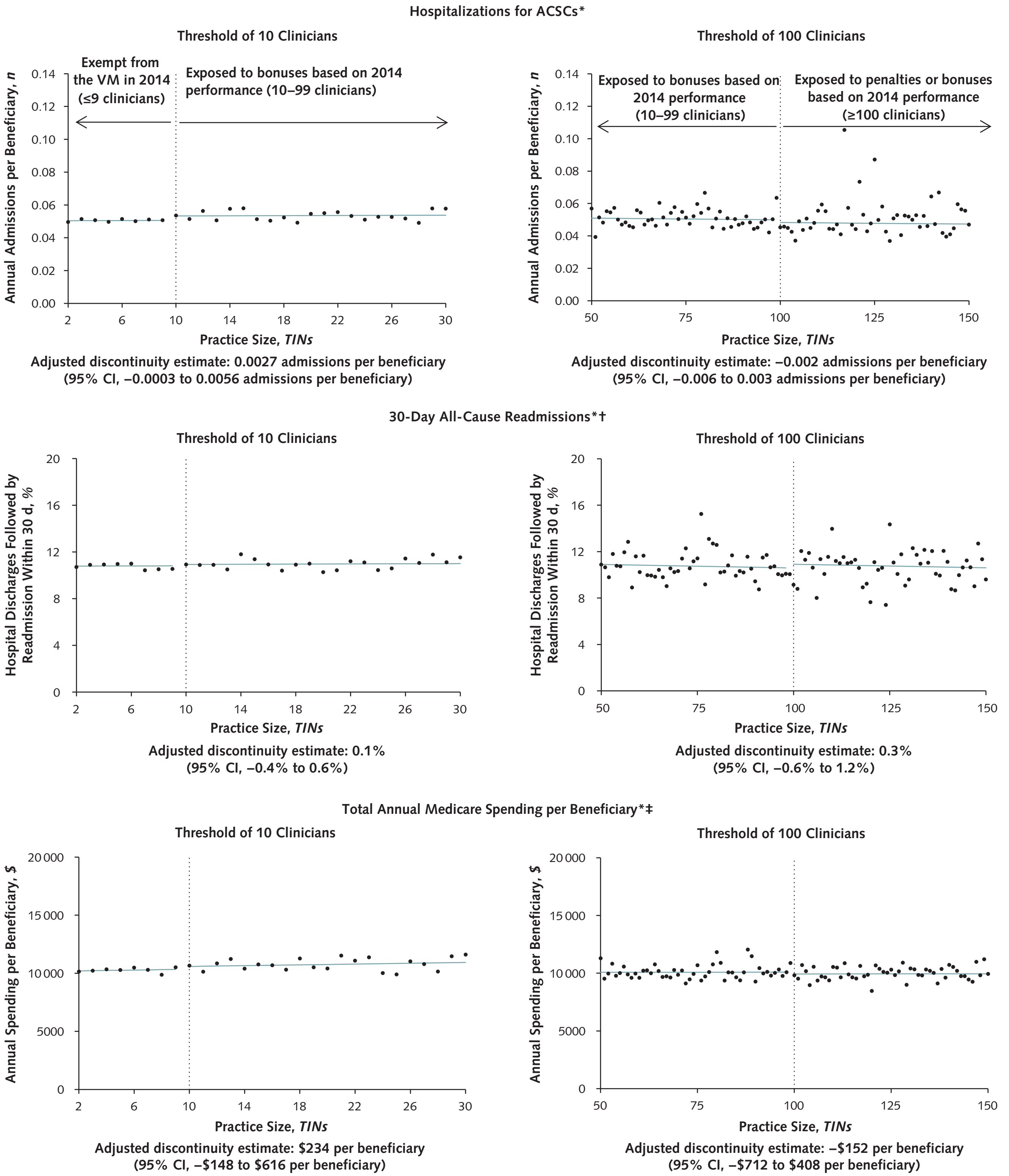
No statistically significant discontinuities were found in the relationship between practice size and patient characteristics at the threshold of 100 or more clinicians (Table 2). We observed discontinuities in a few patient characteristics at the threshold of 10 or more clinicians—notably, higher proportions of beneficiaries with Medicaid coverage and disabilities above the threshold (Supplement Figure 2, available at Annals.org)—but tests conducted at other arbitrary thresholds indicated that these differences were not specific to the 10-clinician threshold and therefore could be attributed to the program as opposed to random chance. We also found no evidence of practice bunching around either threshold (Supplement Figure 3, available at Annals.org).

Analyses of 2012 data produced discontinuity estimates that were similar in magnitude to those from analyses of the 2014 and 2015 data (Supplement Table 2). Results of other sensitivity analyses supported conclusions from our main results.

Effect of Additional Adjustments for Patient Characteristics

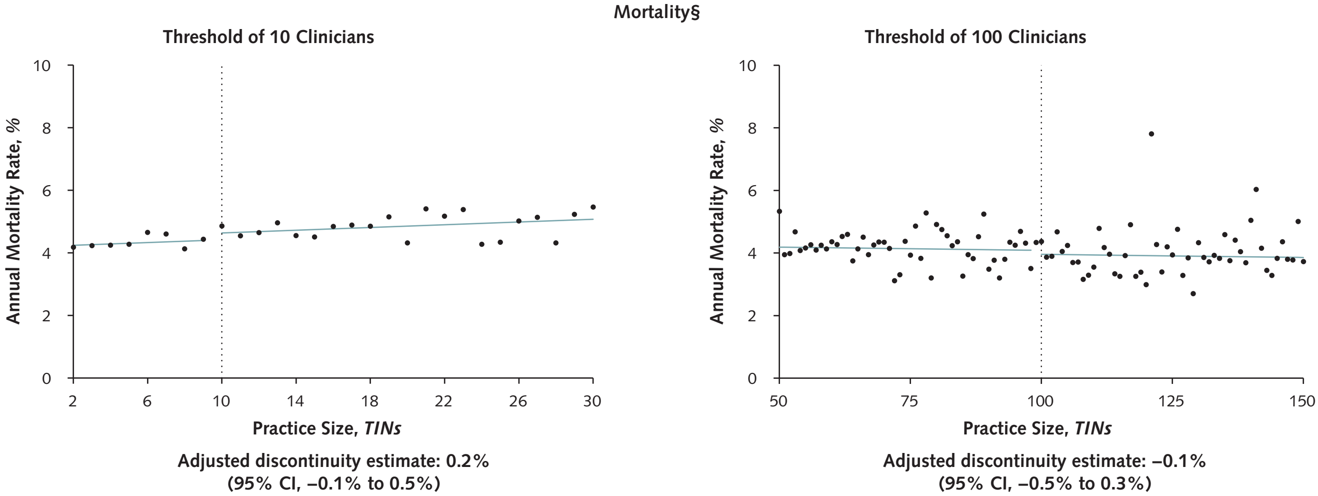
Practices serving disproportionately more patients with dual eligibility or high HCC scores had higher rates of hospitalization for ACSCs, Medicare spending, and mortality (Figure 2) after adjustment for base sets of patient characteristics (Table 1). Additional patient characteristics strongly predicted these outcomes within practices and varied substantially across practices (Supplement Table 5, available at Annals.org). Adjusting for the additional patient characteristics (Table 1) reduced differences between practices in the high-

Figure 1. Discontinuities in the relationship between practice size and performance associated with practice exposure to the VM.



Continued on following page

Figure 1–Continued.



Binned scatter plots of claims-based measures used to assess practice performance for the VM. Each point on the graphs is an average calculated among all patients attributed to practices of a particular size (indicated on the *x*-axis) and is adjusted for the patient characteristics listed in **Table 2**, indicators of 27 Chronic Conditions Data Warehouse chronic conditions, and counts of chronic conditions (**Supplement**). Fitted values from regression discontinuity models, adjusted for patient characteristics and a linear trend in practice size, are superimposed on the scatter plots (green lines). The vertical distance between the fitted lines at the 10- and 100-clinician thresholds (dotted vertical lines) correspond to the regression discontinuity estimates reported below the graphs. The 95% CIs for the regression discontinuity estimates were calculated by using SEs clustered at the practice level. ACSC = ambulatory care–sensitive condition; CMS = Centers for Medicare & Medicaid Services; TIN = taxpayer identification number; VM = Value-Based Payment Modifier.
* Performance measure in the VM.
† Limited to patients with at least 1 index admission during the study year. For each beneficiary, we randomly sampled 1 index admission and assessed whether the patient was readmitted within 30 days of the discharge date of this admission. We excluded admissions in which patients were transferred to another hospital, planned readmissions, and index admissions occurring after December 1 of the year (see **Supplement**, available at Annals.org). Because beneficiaries with several readmissions were sampled only once, the mean readmission rate shown is lower than the annual average for the entire Medicare population. Consistent with CMS methods for risk-adjusting readmissions, our regression discontinuity estimates were also adjusted for 31 indicators of clinical conditions reported on patients' claims in the 365 days preceding the index admission.
‡ Total Medicare Part A and Part B spending per beneficiary (**Supplement**).
§ Assessed from death dates reported (if present) in beneficiary summary files.

est and those in the lowest quartile of the share of dual-eligible patients by 55.9% for hospitalizations for ACSCs, 11.9% for Medicare spending, and 34.8% for mortality (P < 0.001 for all) (**Figure 2**). The additional adjustments reduced differences between practices in the highest quartile of mean HCC score and those in the lowest quartile by 67.9% for hospitalizations for ACSCs, 9.2% for Medicare spending, and 21.6% for mortality (P < 0.001 for all).

Simulations indicated that practice rankings were changed considerably by the additional adjustments, with the most pronounced reordering for hospitalizations for ACSCs, which CMS adjusts only for age and sex (**Table 1**). Practice rankings changed by 1 or more deciles for 61.9% of practices for hospitalizations for ACSCs, 15.3% for Medicare spending, and 30.6% for mortality, with a net movement of poor-performing practices upward and high-performing practices downward in rankings (**Appendix Figure**). The 5% of practices most affected moved ± 27 to 55, ± 5 to 9, and ± 9 to 17 percentiles for hospitalizations for ACSCs, Medicare spending, and mortality, respectively (**Supplement Figure 5**, available at Annals.org).

Under the MIPS, this extent of reordering would be expected to move 2.9% to 16.7% of practices from above to below the exceptional-performance threshold for a given measure and 1.6% to 9.9% of practices from below to above the threshold, depending on the measure (**Supplement Table 7**, available at Annals.org). Under the VM, reordering expected from the additional adjustments would have moved 4.8% to 25.7% of practices out of eligibility for bonuses for a given measure and 3.7% to 24.9% of practices out of eligibility for penalties.

DISCUSSION

Differences in the exposure of physician practices to financial incentives in the VM were not associated with meaningful differences in hospitalization for ACSCs, readmissions, mortality, or Medicare spending after 1 or 2 years of exposure. Several features of the VM may have contributed to this lack of an association. The penalties were modest (a maximum of 2% in 2014 and 4% in 2015) and were applied to Part B payments only. Although bonuses were much larger than penalties, practices had to perform at least 1 SD better than the mean to be eligible for a bonus (10, 37), which may have weakened incentives for poor performers to improve. In addition, some practices may have been unaware of the program, and others may have needed more than 2 years to respond effectively to the incentives, even if they found them strong enough to warrant a response.

Incentives to improve quality and lower spending in the MIPS may be somewhat stronger or weaker than those in the VM but share many features that make them weak overall (21). Under the MIPS, more practices will receive payment adjustments than under the VM, and practices with incrementally higher performance scores will receive proportionally larger bonuses or smaller penalties, thereby strengthening incentives for low performers to improve (1, 38). On the other hand, practices have control over selecting quality measures in the MIPS—and thus have opportunities for gaming that may greatly diminish incentives to improve quality—whereas the VM assessed all practices on core measures (39). As in the VM, incentives in the MIPS to decrease spending are weak, because spending measures are given little weight in overall performance scores. In addition, because bonuses in the VM and MIPS are structured as fee increases, practices receiving bonuses have weaker incentives to limit their provision of Part B services (21).

It therefore is not surprising that VM exposure was not associated with better quality or lower spending, and it would not be surprising if the effect of the MIPS was similarly negligible in its first few years. In contrast to our findings for the VM, stronger incentives to lower spending in the Medicare ACO programs have been associated with significant reductions in Medicare spending within 2 years of program implementation (5, 40–44).

Our findings also suggest that pay-for-performance programs with weak incentives and inadequate risk adjustment might contribute to health care disparities without eliciting a behavioral change that improves care on average. Specifically, we found that adjusting for additional patient characteristics narrowed performance differences between practices serving disproportionately more medically complex and low-income patients and those serving fewer of these patients. Thus, inadequate risk adjustment for clinical and socioeconomic factors may lead to sustained transfers of payments away from practices serving poorer and sicker patients for reasons not related to quality or efficiency of care.

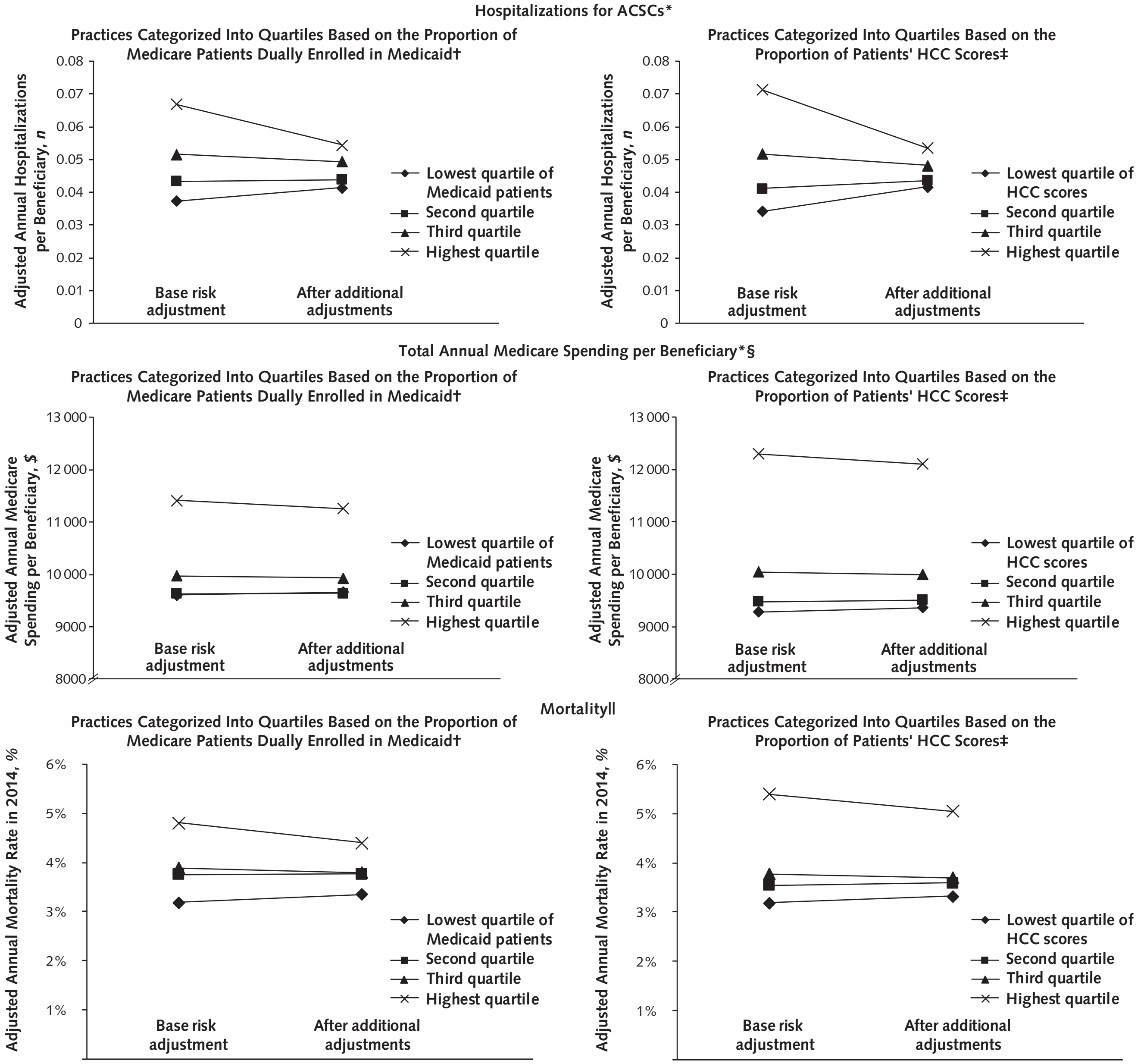
Our study therefore suggests benefits of more complete risk adjustment in the MIPS. It may be impossible or impractical, however, to collect data that capture all relevant differences in patient mix across practices (18, 19). Thus, even enhanced risk adjustment may not fully insulate practices from penalties that reflect patient risk rather than only quality of care. In addition to depleting providers' resources to improve care for vulnerable patients, these penalties might create incentives for practices to avoid sicker or poorer patients. To mitigate these unintended consequences, a portion of

Table 2. Differences in Patient Characteristics Associated With 100-Clinician Practice Size Threshold for Exposure to the VM

Patient Characteristic	Mean (IQR*)			Estimated Discontinuity at 100 Clinicians (95% CI)†
	2–9 Clinicians (1 019 083 Patients‡; 36 250 Practices)	10–99 Clinicians (961 183 Patients‡; 8491 Practices)	≥100 Clinicians (1 095 037 Patients‡; 931 Practices)	
Male, %	42.8 (29.2 to 57.1)	42.8 (36.2 to 53.8)	42.3 (40.0 to 45.6)	0.4 (−0.9 to 1.6)
Mean age, *y*	72.2 (66.3 to 74.5)	71.9 (64.0 to 73.2)	71.7 (66.9 to 72.7)	−0.3 (−1.2 to 0.6)
Race/ethnicity, %				
White	83.6 (70.0 to 99.9)	85.1 (71.4 to 96.6)	83.6 (67.8 to 93.5)	1.2 (−2.4 to 4.8)
Black	8.6 (0 to 10.7)	8.1 (0 to 14.3)	8.9 (1.3 to 15.7)	−1.2 (−4.2 to 1.8)
Hispanic	4.6 (0 to 2.9)	3.8 (0 to 5.1)	3.7 (0.6 to 6.0)	0.01 (−1.05 to 1.07)
Other	3.2 (0 to 2.1)	3.0 (0 to 3.6)	3.7 (1.3 to 4.0)	−0.05 (−1.49 to 1.40)
Enrolled in Medicaid, %§	15.2 (0 to 25.6)	15.6 (5.6 to 33.3)	15.1 (9.5 to 26.7)	−0.4 (−3.7 to 2.8)
Disabled, %‖	22.7 (3.1 to 42.6)	23.8 (16.6 to 50.0)	23.3 (19.3 to 38.7)	0.5 (−2.8 to 3.8)
CCW chronic conditions, *n*¶	6.0 (3.8 to 6.5)	5.9 (4.0 to 6.3)	5.7 (5.1 to 6.0)	−0.05 (−0.28 to 0.19)
End-stage renal disease, %	1.3 (0 to 1.3)	1.2 (0 to 1.4)	1.3 (0.6 to 2.1)	−0.3 (−0.5 to 0)
HCC score**	1.30 (0.77 to 1.45)	1.30 (0.92 to 1.48)	1.28 (1.17 to 1.44)	−0.03 (−0.09 to 0.04)

CCW = Chronic Conditions Data Warehouse; HCC = Hierarchical Condition Category; IQR = interquartile range; VM = Value-Based Payment Modifier.
* Twenty-fifth and 75th percentiles of the practice-level distribution of the characteristic shown.
† Difference in 2014 for patients above versus below the 100-clinician threshold of practice size, adjusted for the linear trend in the characteristic as a function of practice size. 95% CIs were estimated by using SEs clustered at the practice level.
‡ In a random 20% sample of fee-for-service Medicare beneficiaries.
§ Dual-eligible patients included those with full Medicaid enrollment (excluding partial Medicaid enrollees in the Qualified Medicare Beneficiary, Specified Low-Income Medicare Beneficiary, and Qualifying Individual programs).
‖ Disability was the original reason for Medicare eligibility.
¶ Using the Medicare CCW, which draws from Medicare claims since 1999 to characterize each beneficiary's accumulated burden of chronic disease, we assessed the presence of 27 chronic conditions reported before the study year: Alzheimer disease, Alzheimer disease and related disorders or senile dementia, anemia, asthma, atrial fibrillation, benign prostatic hyperplasia, breast cancer, cataract, chronic kidney disease, chronic obstructive pulmonary disease, colorectal cancer, depression, diabetes, endometrial cancer, glaucoma, heart failure, hip or pelvic fracture, hyperlipidemia, hypertension, hypothyroidism, ischemic heart disease, lung cancer, osteoporosis, prostate cancer, acute myocardial infarction, rheumatoid arthritis, and stroke or transient ischemic attack.
** Derived from demographic characteristics and diagnoses in Medicare enrollment and claims files, with higher risk scores indicating higher predicted spending in the subsequent year. For each beneficiary, we constructed the HCC score by using Medicare enrollment and claims data from the previous year.

Figure 2. Risk-adjusted differences between practices serving patients with higher and those serving patients with lower rates of Medicaid enrollment and HCC risk scores, before and after adjustment for additional patient characteristics.



The graphs show the average performance of the first through fourth quartiles of practices, grouped on the basis of the proportion of patients with dual enrollment in Medicaid or of patients' HCC scores, where practice performance is risk-adjusted for the base variables used in CMS risk-adjustment methods and for all patient characteristics listed in Table 1. The proportion of patients dually enrolled in Medicare and Medicaid was 5.3%, 12.7%, 20.9%, and 50.1% in the lowest, second, third, and highest quartiles, respectively. The mean HCC score was 1.00, 1.22, 1.34, and 1.90 in the first, second, third, and fourth quartiles, respectively. For all outcomes, we observed a statistically significant (P < 0.001) reduction in performance differences between the highest and lowest quartiles of practices as a result of the additional adjustments (see Supplement, available at Annals.org, for details). For hospitalization for ACSCs, Medicare spending, and mortality, the additional adjustments narrowed differences between the highest and lowest quartile of practices (grouped by patients' dual-eligibility status) by 55.9%, 11.9%, and 34.8%, respectively. The additional adjustments reduced differences between the highest and lowest quartiles of practices (grouped by patients' HCC scores) by 67.9%, 9.2%, and 21.6%, respectively, for hospitalization for ACSCs, Medicare spending, and mortality. ACSC = ambulatory care-sensitive condition; CMS = Centers for Medicare & Medicaid Services; HCC = Hierarchical Condition Category; VM = Value-Based Payment Modifier.

* Performance measure in the VM.

† In the lowest quartile of practices with dual-eligible patients, the mean Medicaid enrollment rate was 5.3%, and it was 12.7%, 20.9%, and 50.1% in the second, third, and fourth quartiles, respectively.

‡ In the lowest quartile of practices based on HCC scores, the mean practice HCC score was 1.00, and it was 1.22, 1.34, and 1.90 in the second, third, and fourth quartiles, respectively.

§ Total Medicare Part A and Part B spending per beneficiary (Supplement).

‖ Assessed from death dates reported (if present) in beneficiary summary files.

payments to practices in the MIPS might take the form of a per patient monthly payment (such as a care management fee) that is greater for higher-risk patients and does not depend on practice performance, as in the Comprehensive Primary Care Plus model (21, 45). In addition, the extreme tails of spending and utilization measures (in which standard risk-adjustment methods fail most) might be excluded from performance assessments (46, 47). Because the costs to providers of achieving high performance on outcome measures likely are greater for patients with clinical and social risk factors for poor outcomes, our findings suggest that—until such remedies are implemented—programs like the VM and MIPS will impose the costs of serving higher-risk patients on providers, in the form of either penalties for poor performance or higher costs of care improvement (21).

Our study had several limitations. First, we could not measure practices' exposure to the VM directly (6, 9, 10), but our assessment of practice size closely reproduced CMS-reported totals for practices with 100 or more clinicians. Second, although we adjusted for patient characteristics observable in Medicare administrative data, we could not adjust for other risk factors that may have further affected practice performance assessments, such as self-reported health status, functional limitations, education, and cognition (18, 24–26, 48). Third, our analyses of some outcome measures lacked sufficient statistical power to detect small effects of the VM. However, regression discontinuity estimates were not larger than those observed in 2012 or at size thresholds where there were no differences in program exposure, and we found no consistent evidence of growth in effects in 2015.

In conclusion, financial incentives in the VM were not associated with meaningful differences in admissions for ACSCs, readmissions, Medicare spending, or mortality. Performance differences between practices serving higher-risk patients and those serving lower-risk patients were affected considerably by adjustment for additional patient characteristics, highlighting the potential for Medicare's pay-for-performance programs to exacerbate health care disparities and the need for strategies to minimize unintended consequences of these programs for vulnerable populations.

From the Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, and Harvard Medical School, Boston, Massachusetts; Harvard Medical School, Boston, Massachusetts; and Harvard Medical School and Brigham and Women's Hospital, Boston, Massachusetts.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Laura and John Arnold Foundation.

Grant Support: By grants from the Laura and John Arnold Foundation, National Institute on Aging of the National Institutes of Health (P01 AG032952), and Marshall J. Seidman Center for Studies in Health Economics and Health Care Policy at Harvard Medical School.

Disclosures: Dr. McWilliams reports personal fees from Abt Associates and the Medicare Payment Advisory Commission outside the submitted work. Authors not named here have disclosed no conflicts of interest. Disclosures can also be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M17-1740.

Reproducible Research Statement: *Study protocol:* As described in Methods and the Supplement. *Statistical code:* Available from Dr. Roberts (e-mail, eric.roberts@pitt.edu). *Data set:* Available from CMS under its procedures for research data.

Requests for Single Reprints: J. Michael McWilliams, MD, PhD, Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115; e-mail, mcwilliams@hcp.med.harvard.edu.

Current author addresses and author contributions are available at Annals.org.

References

1. Centers for Medicare & Medicaid Services (CMS), HHS. Medicare program; Merit-Based Incentive Payment System (MIPS) and Alternative Payment Model (APM) incentive under the physician fee schedule, and criteria for physician-focused payment models. Final rule with comment period. Fed Regist. 2016;81:77008-831. [PMID: 27905815]
2. Burwell SM. Setting value-based payment goals—HHS efforts to improve U.S. health care. N Engl J Med. 2015;372:897-9. [PMID: 25622024] doi:10.1056/NEJMp1500445
3. VanLare JM, Blum JD, Conway PH. Linking performance with payment: implementing the Physician Value-Based Payment Modifier. JAMA. 2012;308:2089-90. [PMID: 23198298]
4. Anderson GF, Davis K, Guterman S. Medicare payment reform: aligning incentives for better care. Issue Brief (Commonw Fund). 2015;20:1-12. [PMID: 26151988]
5. McWilliams JM, Hatfield LA, Chernew ME, Landon BE, Schwartz AL. Early performance of accountable care organizations in Medicare. N Engl J Med. 2016;374:2357-66. [PMID: 27075832] doi:10.1056/NEJMsa1600142
6. Centers for Medicare & Medicaid Services. Medicare program; revisions to payment policies under the physician fee schedule, clinical laboratory fee schedule, access to identifiable data for the Center for Medicare and Medicaid Innovation Models & other revisions to Part B for CY 2015. Fed Regist. 2014;79:67931-41.
7. Centers for Medicare & Medicaid Services. Timeline to phase in the Value-Based Payment Modifier. Updated 21 September 2017. Accessed at www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Timeline.html on 10 December 2016.
8. Centers for Medicare & Medicaid Services. 2015 Value-Based Payment Modifier Program Experience Report. Baltimore: Centers for Medicare & Medicaid Services; 2015.
9. Centers for Medicare & Medicaid Services (CMS), HHS. Medicare program; revisions to payment policies under the physician fee schedule, clinical laboratory fee schedule & other revisions to Part B for CY 2014. Final rule with comment period. Fed Regist. 2013;78:74229-823. [PMID: 24340356]

10. **Centers for Medicare & Medicaid Services.** Detailed Methodology for the 2016 Value Modifier and the 2014 Quality and Resource Use Report. Baltimore: Centers for Medicare & Medicaid Services; 2015.
11. **Centers for Medicare & Medicaid Services.** Detailed Methodology for the 2017 Value Modifier and the 2014 Quality and Resource Use Report. Baltimore: Centers for Medicare & Medicaid Services; 2017.
12. **Centers for Medicare & Medicaid Services.** Physician Groups Receive Upward, Neutral, or Downward Adjustments to Their Medicare Payments in 2016 Based on Their Performance on Quality and Cost Efficiency Measures. Baltimore: Centers for Medicare & Medicaid Services; 2016.
13. **Lowes R.** Medicare penalties for 5477 groups fund bonuses for 128. Medscape. 2016. Accessed at www.medscape.com/viewarticle/860103 on 8 January 2017.
14. **Centers for Medicare & Medicaid Services.** 2014 Measure Information about the Acute and Chronic Ambulatory Care-Sensitive Condition Composite Measures, Calculated for the Value-Based Payment Modifier Program. Baltimore: Centers for Medicare & Medicaid Services; 2015.
15. **Centers for Medicare & Medicaid Services.** 2014 Measure Information about the Per Capita Costs for All Attributed Beneficiaries Measure, Calculated for the Value-Based Payment Modifier Program. Baltimore: Centers for Medicare & Medicaid Services; 2015.
16. **Centers for Medicare & Medicaid Services.** 2014 Measure Information about the 30-Day All-Cause Hospital Readmission Measure, Calculated for the Value-Based Payment Modifier Program. Baltimore: Centers for Medicare & Medicaid Services; 2015.
17. **U.S. Department of Health and Human Services.** Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Payment Programs. Washington, DC: U.S. Department of Health and Human Services Assistant Secretary for Planning and Evaluation; 2016.
18. **Buntin MB, Ayanian JZ.** Social risk factors and equity in Medicare payment. N Engl J Med. 2017;376:507-10. [PMID: 28177864] doi:10.1056/NEJMp1700081
19. **National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Population Health and Public Health Practice; Board on Health Care Services; Committee on Accounting for Socioeconomic Status in Medicare Payment Programs.** Accounting for Social Risk Factors in Medicare Payment: Data. Washington, DC: National Academies Pr; 2016.
20. **Jha AK, Zaslavsky AM.** Quality reporting that addresses disparities in health care. JAMA. 2014;312:225-6. [PMID: 25027134] doi:10.1001/jama.2014.7204
21. **McWilliams JM.** MACRA: big fix or big problem? Ann Intern Med. 2017;167:122-4. [PMID: 28505630] doi:10.7326/M17-0230
22. **Chen LM, Epstein AM, Orav EJ, Filice CE, Samson LW, Joynt Maddox KE.** Association of practice-level social and medical risk with performance in the Medicare Physician Value-Based Payment Modifier Program. JAMA. 2017;318:453-61. [PMID: 28763549] doi:10.1001/jama.2017.9643
23. **Ryan AM.** Will value-based purchasing increase disparities in care? N Engl J Med. 2013;369:2472-4. [PMID: 24369072] doi:10.1056/NEJMp1312654
24. **Barnett ML, Hsu J, McWilliams JM.** Patient characteristics and differences in hospital readmission rates. JAMA Intern Med. 2015;175:1803-12. [PMID: 26368317] doi:10.1001/jamainternmed.2015.4660
25. **Rose S, Zaslavsky AM, McWilliams JM.** Variation in accountable care organization spending and sensitivity to risk adjustment: implications for benchmarking. Health Aff (Millwood). 2016;35:440-8. [PMID: 26953298] doi:10.1377/hlthaff.2015.1026
26. **Franks P, Fiscella K.** Effect of patient socioeconomic status on physician profiles for prevention, disease management, and diagnostic testing costs. Med Care. 2002;40:717-24. [PMID: 12187185]
27. **Friedberg MW, Safran DG, Coltin K, Dresser M, Schneider EC.** Paying for performance in primary care: potential impact on practices and disparities. Health Aff (Millwood). 2010;29:926-32. [PMID: 20439882] doi:10.1377/hlthaff.2009.0985
28. **Venkataramani AS, Bor J, Jena AB.** Regression discontinuity designs in healthcare research. BMJ. 2016;352:i1216. [PMID: 26977086] doi:10.1136/bmj.i1216
29. **Centers for Medicare & Medicaid Services**. Two-Step Attribution for Measures Included in the Value Modifier. Baltimore: Centers for Medicare & Medicaid Services; 2015.
30. **Centers for Medicare & Medicaid Services; Research Data Assistance Center.** Shared Savings Program Accountable Care Organizations (ACO) provider-level research identifiable files. October 2016. Accessed at www.resdac.org/cms-data/files/ssp-aco-provider-level-rif on 8 January 2017.
31. **Centers for Medicare & Medicaid Services.** Medicare data on provider practice and specialty (MD-PPAS). 2016. Accessed at www.resdac.org/cms-data/files/md-ppas on 20 December 2016.
32. **Peikes D, Ghosh A, Zutshi A, Taylor EF, Anglin G, Converse L, et al.** Evaluation of the Comprehensive Primary Care Initiative: Second Annual Report. Princeton: Mathematica Policy Research; April 2016.
33. **Pope G, Kautter J, Ingber MJ, Sekar R, Newhart C.** Evaluation of the CMS-HCC Risk Adjustment Model. Research Triange Park, NC: RTI International; 2011.
34. **Lee DS, Lemieux T.** Regression discontinuity designs in economics. J Econ Lit. 2010;48:281-355.
35. **Centers for Medicare & Medicaid Services.** Notice of Proposed Rulemaking—Medicare Access and CHIP Reauthorization Act of 2015 Quality Payment Program. Baltimore: Centers for Medicare & Medicaid Services; 2016.
36. **Centers for Medicare & Medicaid Services; U.S. Department of Health and Human Services.** Medicare program; Merit-Based Incentive Payment System (MIPS) and Alternative Payment Model (APM) incentive under the physician fee schedule, and criteria for physician-focused payment models. Fed Regist. 2016;81:77338.
37. **Centers for Medicare & Medicaid Services.** Computation of the 2016 value modifier. 2015. Accessed at www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/2016-VM-Fact-Sheet.pdf on 20 October 2016.
38. **National Association of ACOs**. The ACOs Guide to MACRA. Washington, DC: National Association of ACOs; 2016:25-37.
39. **Centers for Medicare & Medicaid Services; U.S. Department of Health and Human Services.** Executive summary 42 CFR parts 414 and 495 Medicare Program; Merit-based Incentive Payment System (MIPS) and Alternative Payment Model (APM) incentive under the physician fee schedule, and criteria for physician-focused payment models. Baltimore: Centers for Medicare & Medicaid Services; 2016.
40. **Nyweide DJ, Lee W, Cuerdon TT, Pham HH, Cox M, Rajkumar R, et al.** Association of pioneer accountable care organizations vs traditional Medicare fee for service with spending, utilization, and patient experience. JAMA. 2015;313:2152-61. [PMID: 25938875] doi:10.1001/jama.2015.4930
41. **McWilliams JM, Chernew ME, Landon BE, Schwartz AL.** Performance differences in year 1 of pioneer accountable care organizations. N Engl J Med. 2015;372:1927-36. [PMID: 25875195] doi:10.1056/NEJMsa1414929
42. **McWilliams JM.** Changes in Medicare shared savings program savings from 2013 to 2014 [Letter]. JAMA. 2016;316:1711-3. [PMID: 27612292] doi:10.1001/jama.2016.12049
43. **Colla CH, Fisher ES.** Moving forward with accountable care organizations: some answers, more questions. JAMA Intern Med. 2017;177:527-8. [PMID: 28192555] doi:10.1001/jamainternmed.2016.9122
44. **McWilliams JM, Gilstrap LG, Stevenson DG, Chernew ME, Huskamp HA, Grabowski DC.** Changes in postacute care in the Medicare Shared Savings Program. JAMA Intern Med. 2017;177:518-26. [PMID: 28192556] doi:10.1001/jamainternmed.2016.9115
45. **Center for Medicare and Medicaid Innovation.** CPC+ payment methodologies: beneficiary attribution, care management fee, performance-based incentive payment, and payment under the

Medicare physician fee schedule. Version 2. February 2017. Accessed at https://innovation.cms.gov/Files/x/cpcplus-methodology.pdf on 1 May 2017.

46. **Glied S.** How policymakers can foster organizational innovation in health care. Health Affairs Blog. 15 July 2016. Accessed at www.healthaffairs.org/do/10.1377/hblog20160715.055867/full on 10 September 2017.

47. **McWilliams JM, Schwartz AL.** Focusing on high-cost patients—the key to addressing high costs? N Engl J Med. 2017;376:807-9. [PMID: 28249127] doi:10.1056/NEJMp1612779

48. **Kuye IO, Frank RG, McWilliams JM.** Cognition and take-up of subsidized drug benefits by Medicare beneficiaries. JAMA Intern Med. 2013;173:1100-7. [PMID: 23649604] doi:10.1001/jamainternmed.2013.845

Current Author Addresses: Dr. Roberts: Department of Health Policy and Management, University of Pittsburgh Graduate School of Public Health, A653 Crabtree Hall, 130 De Soto Street, Pittsburgh, PA 15261.

Drs. Zaslavsky and McWilliams: Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115.

Author Contributions: Conception and design: E.T. Roberts, A.M. Zaslavsky, J.M. McWilliams.

Analysis and interpretation of the data: E.T. Roberts, A.M. Zaslavsky, J.M. McWilliams.

Drafting of the article: E.T. Roberts, J.M. McWilliams.

Critical revision for important intellectual content: E.T. Roberts, A.M. Zaslavsky, J.M. McWilliams.

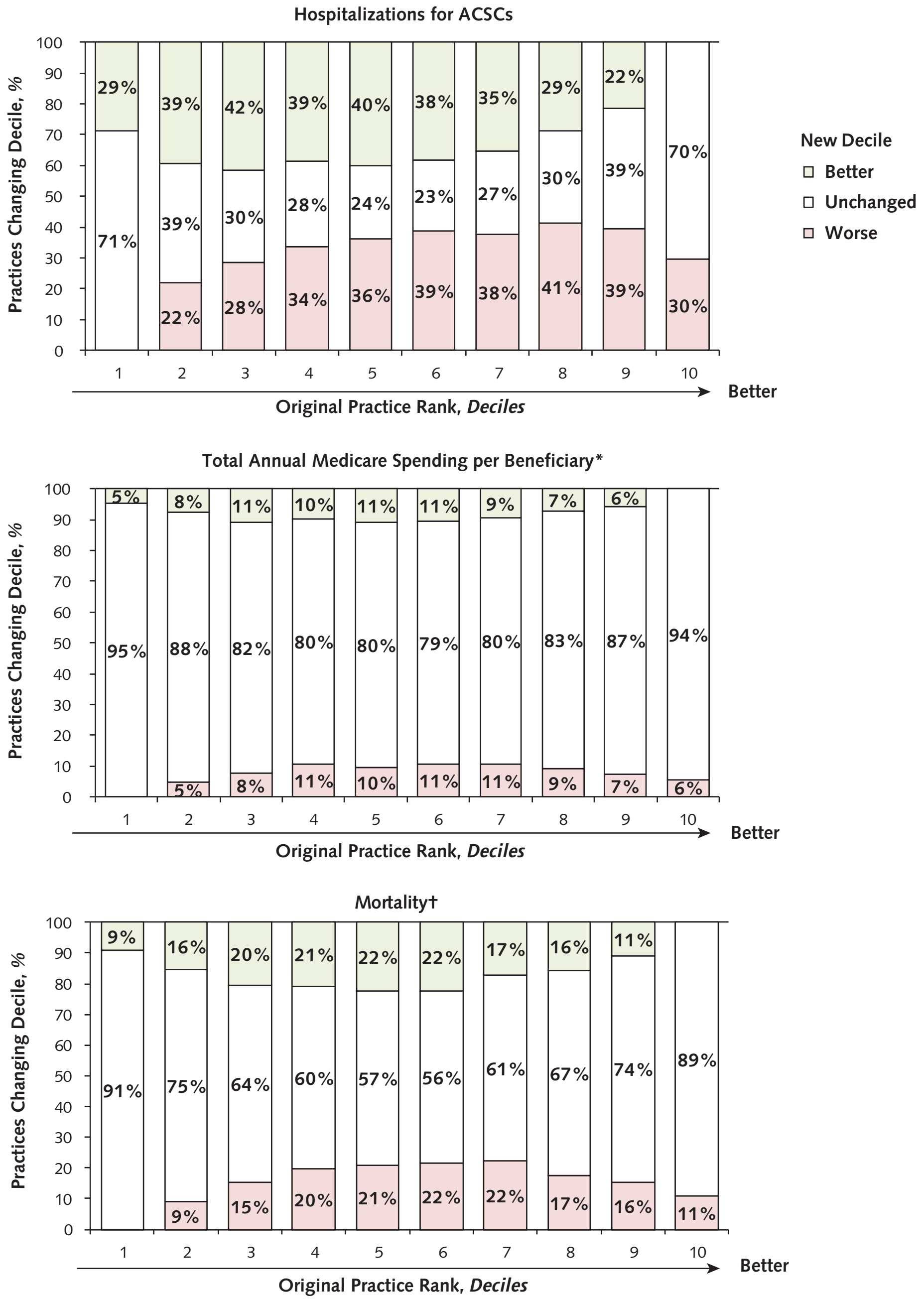
Final approval of the article: E.T. Roberts, A.M. Zaslavsky, J.M. McWilliams.

Statistical expertise: A.M. Zaslavsky, J.M. McWilliams.

Obtaining of funding: J.M. McWilliams.

Collection and assembly of data: E.T. Roberts, J.M. McWilliams.

Appendix Figure. Simulated changes in practice rankings resulting from the use of base CMS risk adjustment versus adjustment for additional patient characteristics.



These graphs summarize practice performance under 2 risk-adjustment approaches: adjustment for the base variables used in CMS risk-adjustment methods and adjustment for the additional patient-level factors listed in **Table 1**. For each outcome, we simulated the proportion of practices whose performance ranking would change by ≥1 decile after additional adjustments. Simulations were based on 10 000 draws from a multivariate normal distribution based on the empirical variances and correlations of practice performance under the 2 risk-adjustment approaches (**Supplement**, available at Annals.org). ACSC = ambulatory care–sensitive condition; CMS = Centers for Medicare & Medicaid Services.
* Total Medicare Part A and Part B spending per beneficiary (**Supplement**).
† Assessed from death dates reported (if present) in beneficiary summary files.