Published in final edited form as:

Ann Intern Med. 2018 February 20; 168(4): 255–265. doi:10.7326/M17-1740.

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

Eric T. Roberts, Ph.D., Alan M. Zaslavsky, Ph.D., and J. Michael McWilliams, M.D., Ph.D.

Department of Health Policy and Management, University of Pittsburgh Graduate School of Public Health (ETR), the Department of Health Care Policy, Harvard Medical School in Boston, MA (ETR, AZ, and JMM), and the Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital in Boston (JMM).

Abstract

Background—When risk adjustment is inadequate and incentives are weak, pay-for-performance programs like Medicare's Value Based Payment Modifier (VM) 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 vs. lower-risk patients.

Design—Exploiting the phase-in of the VM based on practice size, we used regression discontinuity analysis and 2014 Medicare claims to estimate differences in practice performance associated with exposure of practices with 100 clinicians to full VM incentives (bonuses and penalties) and exposure of practices with 10 clinicians to partial incentives (bonuses only). We repeated analyses with 2015 claims to estimate performance differences associated with a second year of exposure above the 100-clinician threshold. We assessed differences in performance between practices serving higher-risk vs. lower-risk patients after standard Medicare adjustments vs. after 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—There were no statistically significant discontinuities at the 10 or (100-clinician thresholds in the relationship between practice size and performance on quality or spending measures in either year. Adjustment for additional patient characteristics narrowed performance

Address correspondence to: J. Michael McWilliams, M.D., Ph.D., Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave., Boston, MA 02115; phone: 617-432-3290; fax: 617-432-0173; mcwilliams@hcp.med.harvard.edu.
Eric T. Roberts, Ph.D., Department of Health Policy & Management, University of Pittsburgh Graduate School of Public Health, A653 Crabtree Hall, 130 De Soto St., Pittsburgh, PA 15261
Alan M. Zaslavsky, Ph.D., Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave., Boston, MA 02115
J. Michael McWilliams, M.D., Ph.D. (corresponding author), Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave., Boston, MA 02115

differences by 9.2–67.9% between practices in the highest vs. lowest quartile of Medicaid patients and Hierarchical Condition Category scores.

Limitations—Observational design, administrative data.

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

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

In January 2017, 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 lowering spending.

Although the effects of the MIPS will not be known for several years, its basic design is similar to that of its predecessor, the Value-Based Payment Modifier (VM).(6) In 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 two years later based on practices' performance scores.(7)

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

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 could result in sustained and unwarranted transfers of resources from practices serving sicker or more socially disadvantaged patients to practices serving healthier or more affluent patients. (23–27)

In evaluating the merits of pay-for-performance programs, it is therefore 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 and 100 clinicians to partial or full VM incentives in 2014, respectively, and performance differences associated with the exposure of practices with 100 clinicians to a second year of incentives in 2015. In a second set of analyses, we examined the impact of adjusting for additional patient characteristics on practice rankings and on performance differences between practices with higher vs. lower proportions of low-income and medically complex 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 vs. below the size thresholds determining exposure to the VM. This design exploits the fact that practice exposure to performance incentives in the VM differed above from below specific thresholds, but that other determinants of spending and quality likely did not, making possible an inference similar to that from a randomized study.⁽²⁸⁾ Because there may be too few observations 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 (i.e., level shifts) in outcomes above vs. below a threshold. Thus, we analyzed data from practices with 50–150 clinicians (for the 10-clinician threshold) and 2–30 clinicians (for 10-clinician thresholds) to estimate discontinuities in spending and quality and assumed that the relationship between practice size and performance would have trended approximately linearly across these thresholds without interruption in the absence of the VM.

In our second set of analyses, we assessed practice performance before vs. after adjusting for additional patient characteristics not included in the risk-adjustment methods used by the Centers for Medicare and Medicaid Services (CMS) in the VM or the MIPS.^(14,15) We compared performance differences between practices serving higher vs. lower proportions of low-income and medically complex patients and compared these differences before and after the additional adjustments. We also assessed changes in the relative 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 Part A and B of fee-for-service Medicare in the year of interest (while alive in the case of decedents) and the prior 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 office visits for the beneficiary during

the study year (Appendix).(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 (Appendix).(30)

Practice Exposure to the VM

To determine practice (TIN) size and thus exposure to VM incentives, we used the 2014 Medicare Provider Practice and Specialty (MD-PPAS) file to attribute each clinician to the TIN(s) under which they billed for Part B services (Appendix).(31) We calculated the total number of clinicians billing under each TIN and created indicators for 3 size categories, each with different exposure to VM incentives in 2014: <10 clinicians (no exposure), 10–99 clinicians (exposed to potential bonuses only), and ≥100 clinicians (exposed to potential bonuses and penalties). Our method for determining practice size closely followed the approach used by CMS to determine practice size for the VM and yielded a number of practices with ≥100 clinicians that was very similar to the number reported by CMS (Appendix Table 1).(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 (Appendix).(16) Although not included as a performance measure in the VM, we also assessed annual mortality as an additional measure that may be particularly sensitive to risk adjustment and can be interpreted more reliably as a health outcome than utilization-based quality measures (e.g., admissions and readmissions may often be appropriate and improve health).

Patient Characteristics

We used Medicare enrollment data to determine beneficiaries' age, sex, race and ethnicity, presence of end-stage renal disease, enrollment in Medicaid (dual eligibility), and whether disability was the original reason for Medicare entitlement. We used the Chronic Conditions Data Warehouse (CCW) to determine the presence of 27 chronic conditions prior to 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 Analyses

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 ≥100-clinician threshold (vs. only bonuses below) and to bonuses above the ≥10-clinician threshold (vs. neither below). Specifically, we fitted a patient-level linear regression model to estimate the difference in performance between practices above vs. below each threshold, adjusting for the linear relationship between the outcome and practice size (number of

clinicians) and for patients' clinical and sociodemographic characteristics (Appendix). This adjusted difference (or adjusted discontinuity) may be interpreted as the difference in performance attributable to VM incentives. We repeated analysis of the 100-clinician threshold using data from 2015, when practices with 10–99 clinicians were additionally exposed to penalties, to isolate performance differences associated with 2 years of exposure to full VM incentives (above the threshold) vs. 1 year of exposure to full incentives (below the threshold).

We conducted several analyses to test the assumptions of our regression discontinuity approach and 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 (Appendix).(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 95–105 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 50–150 clinicians in 2014 moved above or below the 100-clinician threshold 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 and/or low income) and separately based on the mean HCC score of patients in each practice. We estimated differences in performance between quartiles of practices first after 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 perform adjustments, we first used 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 based on their patients' characteristics, ignoring the practice's distinct contribution to quality or spending (Appendix). We then adjusted estimates for each practice by subtracting expected performance from observed performance. If, for example, high-risk patients experienced 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 uses a continuous scoring system to rank practices based on their performance relative to other practices reporting on the same measures and determines bonuses and penalties from the rankings (Appendix).⁽³⁵⁾ Based on 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 the most affected 5% of practices. Finally, we estimated the proportion of practices that would have experienced changes in eligibility for VM bonuses or penalties (those performing ± 1 standard deviation above or below the mean) from the additional adjustments (Appendix).

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 based on the full composite quality scores calculated by the programs. Rather, we could only do so for each claims-based measure we analyzed. Assessments for individual measures are nevertheless instructive because the expected impact 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 in classifying practices into quartiles based on dual eligibility and HCC scores, we focused our second set of analyses on practices with ≥ 100 clinicians and excluded readmissions as an outcome, as 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 vs. below the size thresholds (the adjusted discontinuities) were small in 2014 and not statistically significant (Figure 1). For example, exceeding the ≥ 10 -clinician threshold was associated with an average of 0.003 more hospitalizations for ACSCs per beneficiary (95% CI: $-0.0003, 0.0056$; $P=0.078$), and exceeding the ≥ 100 clinician threshold was associated with an average of 0.002 fewer hospitalizations for ACSCs per beneficiary (95% CI: $0.006, 0.003$; $P=0.48$) than expected from the linear relationship between admission rates and practice size. Analyses of the ≥ 100 -clinician threshold using 2015 data revealed no statistically significant discontinuities associated with a second year of full exposure to the VM (Appendix Table 2).

There were no statistically significant discontinuities in the relationship between practice size and patient characteristics at the ≥ 100 -clinician threshold (Table 2). We observed discontinuities in a few patient characteristics at the ≥ 10 -clinician threshold—notably, higher proportions of beneficiaries with Medicaid coverage and disabilities above the

threshold (Appendix Figure 2)—but tests conducted at other arbitrary thresholds indicated that these differences were not specific to the 10-clinician threshold and therefore not suggestive of program effects as opposed to random chance. We also found no evidence of practice bunching around either threshold (Appendix Figure 3).

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

Impact 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 adjusting for base sets of patient characteristics (Table 1). Additional patient characteristics were strongly predictive of these outcomes within practices and varied substantially across practices (Appendix Table 5). Adjusting for the additional patient characteristics (Table 1) reduced differences between practices in the highest vs. lowest quartile of the dual-eligible share of patients by 55.9% for hospitalization 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 vs. lowest quartile of mean HCC score by 67.9% for hospitalization 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 re-ordering for ACSCs, which CMS adjusts only for age and sex (Table 1). Practice rankings changed by ± 1 decile for 61.9% of practices for hospitalization for ACSCs, 15.3% of practices for Medicare spending, and 30.6% for mortality, with a net movement of poor-performing practices upward in rankings and high-performing practices downward (Figure 3). The most affected 5% of practices moved ± 27 –55, ± 5 –9, and ± 9 –17 percentiles for hospitalization for ACSCs, Medicare spending, and mortality, respectively (Appendix Figure 5).

Under the MIPS, this extent of re-ordering would be expected to move 2.9–16.7% of practices from above to below the exceptional performance threshold for a given measure and 1.6–9.9% of practices from below to above it, depending on the measure (Appendix Table 7). Under the VM, reordering expected from the additional adjustments would have moved 4.8–25.7% of practices out of eligibility for bonuses for a given measure and 3.7–24.9% 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 could 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 one

standard deviation 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 two years to respond effectively to the incentives even if they found them sufficiently strong to warrant a response.

Incentives to improve quality and lower spending in the MIPS may be somewhat stronger or weaker than 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 selection of 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 provision of Part B services.(21)

It is therefore not surprising that VM exposure was not associated with better quality or lower spending and would not be surprising if the impact 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 two years of program implementation.(5,40–44)

Our findings also suggest that pay-for-performance programs with weak incentives and inadequate risk adjustment could 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 vs. fewer medically complex and low-income patients. Thus, inadequate risk adjustment for clinical and socioeconomic factors could lead to sustained transfers of payments away from practices serving poorer and sicker patients for reasons not related to practices' 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 captures 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 quality of care. In addition to depleting providers' resources to improve care for vulnerable patients, these penalties could create incentives for practices to avoid the sicker or poorer patients. To mitigate these unintended consequences, a portion of payments to practices in the MIPS could take the form of a per-patient monthly payment (e.g., a care management fee) that is higher for higher-risk patients and independent of practice performance, as in the Comprehensive Primary Care Plus model.(21,45) In addition, the extreme tails of spending and utilization measures (where standard risk-adjustment methods fail most) could be excluded from performance assessments.(46,47) Because the costs to providers of achieving

high performance on outcome measures are likely 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 from poor performance or higher costs of care improvement.⁽²¹⁾

Our study had several limitations. First, we could not measure practice exposure to the VM directly,^(6,9,10) but our assessment of practice size closely reproduced CMS-reported totals for practices with 100 clinicians. Second, while 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 bigger 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 significant differences in admissions for ACSCs, readmissions, Medicare spending, or mortality. Performance differences between practices serving higher- vs. 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by grants from the Laura and John Arnold Foundation, the National Institutes of Health (P01 AG032952), and the Marshall J. Seidman Center for Studies in Health Economics and Health Care Policy at Harvard Medical School. 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.

References

1. 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. Federal register. 2016; 81(214):77008–831. [PubMed: 27905815]
2. Burwell SM. Setting Value-Based Payment Goals — HHS Efforts to Improve U.S. Health Care. *New England Journal of Medicine*. 2015; 372(10):897–9. [PubMed: 25622024]
3. VanLare JM, Blum JD, Conway PH. Linking performance with payment: Implementing the physician value-based payment modifier. *JAMA*. 2012; 308(20):2089–90. [PubMed: 23198298]
4. Anderson GF, Davis K, Guterman S. Medicare Payment Reform: Aligning Incentives for Better Care. Issue brief. 2015; 20:1–12. [PubMed: 26151988]

5. McWilliams JM, Hatfield LA, Chernew ME, Landon BE, Schwartz AL. Early Performance of Accountable Care Organizations in Medicare. *The New England journal of medicine*. 2016; 374(24):2357–66. [PubMed: 27075832]
6. 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. 2014; 79(219):67931–41.
7. Centers for Medicare and Medicaid Services. [on December 10, 2016] Timeline to Phase In the Value-Based Payment Modifier. Accessed at <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Timeline.html>
8. 2015 Value-Based Payment Modifier Program Experience Report. Baltimore: Centers for Medicare and Medicaid Services; 2015.
9. 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. *Federal register*. 2013; 78(237):74229–823. [PubMed: 24340356]
10. Detailed Methodology for the 2016 Value Modifier and the 2014 Quality and Resource Use Report. Baltimore, MD: Centers for Medicare and Medicaid Services; 2015.
11. Detailed Methodology for the 2017 Value Modifier and the 2014 Quality and Resource Use Report. Baltimore, MD: Centers for Medicare and Medicaid Services; 2017.
12. 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 and Medicaid Services; 2016.
13. Lowes, R. [on January 8, 2017] Medicare Penalties for 5477 Groups Fund Bonuses for 128. *Medscape*. 2016. Accessed at <http://www.medscape.com/viewarticle/860103>
14. 2014 Measure Information about the Acute and Chronic Ambulatory Care-Sensitive Condition Composite Measures, Calculated for the Value-Based Payment Modifier Program. Baltimore, MD: Centers for Medicare and Medicaid Services; 2015.
15. 2014 Measure Information about the Per Capita Costs for All Attributed Beneficiaries Measure, Calculated for the Value-Based Payment Modifier Program. Baltimore, MD: Centers for Medicare and Medicaid Services; 2015.
16. 2014 Measure Information about the 30-Day All-Cause Hospital Readmission Measure, Calculated for the Value-Based Payment Modifier Program. Baltimore, MD: Centers for Medicare and Medicaid Services; 2015.
17. Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Payment Programs. Washington, DC: United States 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. *New England Journal of Medicine*. 2017; 376(6):507–10. [PubMed: 28177864]
19. Accounting for Social Risk Factors in Medicare Payment: Data. Washington, DC: National Academy of Sciences; 2016.
20. Jha AK, Zaslavsky AM. Quality reporting that addresses disparities in health care. *JAMA*. 2014; 312(3):225–6. [PubMed: 25027134]
21. McWilliams JM. MACRA—Big Fix or Big Problem. *Annals of internal medicine*. 2017
22. Chen LM, Epstein AM, Orav E, Filice CE, Samson L, 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(5):453–61. [PubMed: 28763549]
23. Ryan AM. Will Value-Based Purchasing Increase Disparities in Care? *New England Journal of Medicine*. 2013; 369(26):2472–4. [PubMed: 24369072]
24. Barnett ML, Hsu J, McWilliams J. Patient characteristics and differences in hospital readmission rates. *JAMA internal medicine*. 2015; 175(11):1803–12. [PubMed: 26368317]
25. Rose S, Zaslavsky AM, McWilliams JM. Variation In Accountable Care Organization Spending And Sensitivity To Risk Adjustment: Implications For Benchmarking. *Health affairs*. 2016; 35(3): 440–8. [PubMed: 26953298]

26. Franks P, Fiscella K. Effect of Patient Socioeconomic Status on Physician Profiles for Prevention, Disease Management, and Diagnostic Testing Costs. *Medical care*. 2002; 40(8):717–24. [PubMed: 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 affairs*. 2010; 29(5):926–32. [PubMed: 20439882]
28. Venkataramani AS, Bor J, Jena AB. Regression discontinuity designs in healthcare research. *BMJ*. 2016; 352
29. Two-Step Attribution for Measures Included in the Value Modifier. Baltimore, MD: Centers for Medicare and Medicaid Services; 2015.
30. Shared Savings Program Accountable Care Organizations (ACO) Provider-level Research Identifiable Files. Research Data Assistance Center. Oct.2016
31. Medicare Data on Provider Practice and Specialty (MD-PPAS). Centers for Medicare and Medicaid Services. 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, NJ: Mathematica Policy Research; Apr 13. 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. *Journal of Economic Literature*. 2010; 48(2):281–355.
35. Centers for Medicare and Medicaid Services. Notice of Proposed Rulemaking--Medicare Access and CHIP Reauthorization Act of 2015 Quality Payment Program. 2016
36. 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. *Federal register*. 2016; 81(214):77338.
37. Computation of the 2016 Value Modifier. [on October 20, 2016] Centers for Medicare and Medicaid Services. 2015. Accessed at <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/2016-VM-Fact-Sheet.pdf>
38. The ACOs Guide to MACRA. National Association of ACOs; 2016. p. 25-37.
39. 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 and Medicaid Services; 2016.
40. Nyweide DJ, Lee W, Cuerdon TT, et al. Association of pioneer accountable care organizations vs traditional medicare fee for service with spending, utilization, and patient experience. *JAMA*. 2015; 313(21):2152–61. [PubMed: 25938875]
41. McWilliams JM, Chernew ME, Landon BE, Schwartz AL. Performance Differences in Year 1 of Pioneer Accountable Care Organizations. *New England Journal of Medicine*. 2015; 372(20):1927–36. [PubMed: 25875195]
42. McWilliams JM. Changes in Medicare Shared Savings Program Savings From 2013 to 2014. *JAMA*. 2016; 316(16):1711–3. [PubMed: 27612292]
43. Colla CH, Fisher ES. Moving forward with accountable care organizations: Some answers, more questions. *JAMA internal medicine*. 2017; 177(4):527–8. [PubMed: 28192555]
44. McWilliams J, Gilstrap LG, Stevenson DG, Chernew ME, Huskamp HA, Grabowski DC. Changes in postacute care in the medicare shared savings program. *JAMA internal medicine*. 2017; 177(4): 518–26. [PubMed: 28192556]
45. [Accessed on May 1, 2017] CPC+ Payment Methodologies: Beneficiary attribution, care management fee, performance-based incentive payment, and payment under the Medicare physician fee schedule. Center for Medicare and Medicaid Innovation. 2017. at <https://innovation.cms.gov/Files/x/cpcplus-methodology.pdf>
46. Glied S. How Policymakers Can Foster Organizational Innovation In Health Care. *Health Affairs Blog*. 2016

47. McWilliams JM, Schwartz AL. Focusing on High-Cost Patients — The Key to Addressing High Costs? *New England Journal of Medicine*. 2017; 376(9):807–9. [PubMed: 28249127]
48. Kuye IO, Frank RG, McWilliams JM. Cognition and take-up of subsidized drug benefits by Medicare beneficiaries. *JAMA internal medicine*. 2013; 173(12):1100–7. [PubMed: 23649604]

A. Hospitalizations for Ambulatory Care-Sensitive Conditions^b
Threshold of 10 clinicians

Adjusted discontinuity estimate: +0.0027 admissions/beneficiary
(95% CI: -0.0003 to 0.0056)

Threshold of 100 clinicians

Adjusted discontinuity estimate: -0.002 admissions/beneficiary
(95% CI: -0.006 to 0.003)

B. 30-Day All-Cause Readmissions^{b,c}
Threshold of 10 clinicians

Adjusted discontinuity estimate: +0.1%
(95% CI: -0.4% to 0.6%)

Threshold of 100 clinicians

Adjusted discontinuity estimate: +0.3%
(95% CI: -0.6% to 1.2%)

C. Total annual Medicare spending per Beneficiary^{b,d} **D. Mortality^e**

Figure 1. Discontinuities in the relationship between practice size and performance associated with practice exposure to the Value-Based Payment Modifier

This figure presents binned scatterplots of claims-based measures used to assess practice performance for the Value-Based Payment Modifier. Each point in the graphs is an average calculated among all patients attributed to practices of a particular size (indicated on the horizontal axis), and is adjusted for the patient characteristics listed in Table 2, indicators of 27 Chronic Conditions Data Warehouse (CCW) chronic conditions, and counts of chronic conditions (Appendix). Fitted values from regression discontinuity models, adjusted for patient characteristics and a linear trend in practice size, are superimposed on the scatterplots (blue lines). The vertical distance between the fitted lines at the 10 and 100 clinician thresholds (dashed vertical lines) correspond to the regression discontinuity estimates reported below the graphs. 95% confidence intervals for the regression discontinuity estimates were calculated using standard errors clustered at the practice level.

A. Hospitalizations for Ambulatory Care-Sensitive Conditions ^a **B. Total annual Medicare spending per Beneficiary ^{a,d}** **C. Mortality ^e**

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

The graphs show the average performance of the first through fourth quartiles of practices, grouped based on the proportion of patients with dual enrollment in Medicaid or on patients' on Hierarchical Condition Category (HCC) scores, where practice performance is risk-adjusted: (1) for the base variables used in CMS risk-adjustment methods, and (2) 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 Appendix 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 quartile of practices (grouped by patients' HCC scores) by 67.9%, 9.2%, and 21.6%, respectively, for hospitalization for ACSCs, Medicare spending, and mortality.

Hospitalizations for Ambulatory Care-Sensitive Conditions **Total annual Medicare spending per Beneficiary^b**

Figure 3. Simulated changes in practice rankings using base CMS risk adjustment versus adjustment for additional patient characteristics

These graphs summarize practice performance under two risk-adjustment approaches: adjustment for the base variables used in CMS risk-adjustment methods vs. adjustment for 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 two risk-adjustment approaches (Appendix).

Table 1

Patient characteristics used to risk-adjust practice performance

	Outcome		
	Hospitalizations for Ambulatory-Care-Sensitive Conditions ^a	Total annual Medicare spending per Beneficiary ^a	Mortality
Base risk adjustment ^b	<ul style="list-style-type: none"> ■ Age-by-sex categories 	<ul style="list-style-type: none"> ■ Age ■ Sex ■ Hierarchical Condition Category (HCC) score (includes indicators of dual enrollment in Medicare and Medicaid and disability status) and HCC score squared ■ End-stage renal disease 	<ul style="list-style-type: none"> ■ Age ■ Sex ■ HCC score (includes indicators of dual enrollment in Medicare and Medicaid and disability status) and HCC score squared ■ End-stage renal disease
Additional patient characteristics	<ul style="list-style-type: none"> ■ 70 HCC indicators ^c ■ Chronic Conditions Data Warehouse (CCW) conditions ^d ■ End-stage renal disease ■ Disability as original reason for Medicare enrollment ■ Dual Medicare and Medicaid enrollment ^e ■ Recipients of Medicare Savings program ^f ■ Interactions among variables ^g 	<ul style="list-style-type: none"> ■ CCW conditions ^d ■ Recipients of Medicare Savings program ^f ■ Interactions among variables ^g 	<ul style="list-style-type: none"> ■ CCW conditions ^d ■ Recipients of Medicare Savings program ^f ■ Interactions among variables ^g

^aPerformance measure in the Value-Based Payment Modifier.

^bFor hospital admissions for Ambulatory Care-Sensitive Conditions 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.

^cIndicators for 70 hierarchical condition categories included in the CMS-Hierarchical Condition Category (HCC) risk adjustment model. These indicators are used to construct patients' HCC scores.

^dIndicators for the presence of 27 chronic conditions reported prior to the study year, plus counts of chronic conditions (indicator variables in unit increments from 2–8 conditions and for 9 conditions).

^eIncludes low-income persons under age 65 who qualified for Medicare because of a disability and individuals age 65 who additionally qualified for Medicaid because of low income.

^fIncludes partial Medicaid enrollees in the Qualified Medicare Beneficiary, Specified Low-Income Medicare Beneficiary, and Qualifying Individual programs.

^gTwo-way interactions between: HCC score, count of prior-year Chronic Conditions Data Warehouse (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.

Table 2

Differences in patient characteristics associated with 100 clinician practice size threshold for exposure to the Value-based Payment Modifier

Patient Characteristics	Practice Size:			Discontinuity (difference above vs. below threshold adjusted for linear trend with size) ^a :	
	2–9 Clinicians (Patient N=1,019,083 ^b Practice N=36,250)	10–99 Clinicians (Patient N=961,183 ^b Practice N=8,491)	100 Clinicians (Patient N=1,095,037 ^b Practice N=931)	At 100 Clinicians	
	Mean (Interquartile range ^c)			Estimate	95% CI ^d
Male, %	42.8 (29.2 – 57.1)	42.8 (36.2 – 53.8)	42.3 (40.0 – 45.6)	0.4	(–0.9, 1.6)
Age, years	72.2 (66.3 – 74.5)	71.9 (64.0 – 73.2)	71.7 (66.9 – 72.7)	–0.3	(–1.2, 0.6)
Race/Ethnicity, %					
White	83.6 (70.0 – 99.9)	85.1 (71.4 – 96.6)	83.6 (67.8 – 93.5)	1.2	(–2.4, 4.8)
Black	8.6 (0.0 – 10.7)	8.1 (0.0 – 14.3)	8.9 (1.3 – 15.7)	–1.2	(–4.2, 1.8)
Hispanic	4.6 (0.0 – 2.9)	3.8 (0.0 – 5.1)	3.7 (0.6 – 6.0)	0.01	(–1.05, 1.07)
Other	3.2 (0.0 – 2.1)	3.0 (0.0 – 3.6)	3.7 (1.3 – 4.0)	–0.05	(–1.49, 1.40)
Enrolled in Medicaid, % ^e	15.2 (0.0 – 25.6)	15.6 (5.6 – 33.3)	15.1 (9.5 – 26.7)	–0.4	(–3.7, 2.8)
Disabled, % ^f	22.7 (3.1 – 42.6)	23.8 (16.6 – 50.0)	23.3 (19.3 – 38.7)	0.5	(–2.8, 3.8)
CCW Chronic Conditions, No. ^g	6.0 (3.8 – 6.5)	5.9 (4.0 – 6.3)	5.7 (5.1 – 6.0)	–0.05	(–0.28, 0.19)
End-Stage Renal Disease, %	1.3 (0.0 – 1.3)	1.2 (0.0 – 1.4)	1.3 (0.6 – 2.1)	–0.3	(–0.5, 0.0)
HCC Score ^h	1.30 (0.77 – 1.45)	1.30 (0.92 – 1.48)	1.28 (1.17 – 1.44)	–0.03	(–0.09, 0.04)

^aDifference 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.

^bPatients in a random 20% sample of fee-for-service Medicare beneficiaries.

^c25th and 75th percentiles of the practice-level distribution of the characteristic shown.

^dThe 95% confidence intervals were estimated using standard errors clustered at the practice level.

^eDual 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).

^fDisability was original reason for Medicare eligibility.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

^gUsing the Medicare Chronic Conditions Data Warehouse (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 prior to the study year: Alzheimer's disease, Alzheimer's 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.

^hHierarchical Condition Category (HCC) risk scores are derived from demographic 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 using Medicare enrollment and claims data from the prior year.