Changes in Patient Experiences and Assessment of Gaming Among Large Clinician Practices in Precursors of the Merit-Based Incentive Payment System

Eric T. Roberts, PhD; Zirui Song, MD, PhD; Lin Ding, PhD; J. Michael McWilliams, MD, PhD

Abstract

IMPORTANCE Medicare's Merit-Based Incentive Payment System (MIPS), a public reporting and pay-for-performance program, adjusts clinician payments based on publicly reported measures that are chosen primarily by clinicians or their practices. However, measure selection raises concerns that practices could earn bonuses or avoid penalties by selecting measures on which they already perform well, rather than by improving care—a form of gaming. This has prompted calls for mandatory reporting on a smaller set of measures including patient experiences.

OBJECTIVE To examine (1) practices’ selection of Consumer Assessment of Healthcare Providers and Systems (CAHPS) patient experience measures for quality scoring under the pay-for-performance program and (2) the association between mandated public reporting on CAHPS measures and performance on those measures within precursor programs of the MIPS.

DESIGN, SETTING, AND PARTICIPANTS This cross-sectional study included 2 analyses. The first analysis examined the association between the baseline CAHPS scores of large practices (≥100 clinicians) and practices’ selection of these measures for quality scoring under a pay-for-performance program up to 2 years later. The second analysis examined changes in patient experiences associated with a requirement that large practices publicly report CAHPS measures starting in 2014. A difference-in-differences analysis of 2012 to 2017 fee-for-service Medicare CAHPS data was conducted to compare changes in patient experiences between large practices (111-150 clinicians) that became subject to this reporting mandate and smaller unaffected practices (50-89 clinicians). Analyses were conducted between October 1, 2020, and July 30, 2021.

MAIN OUTCOMES AND MEASURES The primary outcomes of the 2 analyses were (1) the association of baseline CAHPS scores of large practices with those practices’ selection of those measures for quality scoring under a pay-for-performance program; and (2) changes in patient experiences associated with a requirement that large practices publicly report CAHPS measures starting in 2014.

RESULTS Among 301 large practices that publicly reported patient experience measures, the mean (IQR) age of patients at baseline was 71.6 (70.4–73.2) years, and 55.8% of patients were women (IQR, 54.3%–57.7%). Large practices in the top vs bottom quintile of patient experience scores at baseline were more likely to voluntarily include these scores in the pay-for-performance program 2 years later (96.3% vs 67.9%), a difference of 28.4 percentage points (95% CI, 9.4–47.5 percentage points; P = .004). After 2 to 3 years of the reporting mandate, patient experiences did not differentially improve in affected vs unaffected practices (difference-in-differences estimate: −0.03 practice-level standard deviations of the composite score; 95% CI, −0.64 to 0.58; P = .92).

CONCLUSIONS AND RELEVANCE In this cross-sectional study of US physician practices that participated in precursors of the MIPS, large practices were found to select measures on which they

(continued)
were already performing well for a pay-for-performance program, consistent with gaming. However, mandating public reporting was not associated with improved patient experiences. These findings support recommendations to end optional measures in the MIPS but also suggest that public reporting on mandated measures may not improve care.

Introduction

The Medicare Merit-Based Incentive Payment System (MIPS) is the largest public reporting and pay-for-performance program for clinicians. The MIPS was introduced in 2017 and scores approximately 880,000 clinicians on 4 domains of care (quality, clinical practice improvement, use of interoperable health information technology, and cost), which are publicly reported on Medicare’s Physician Compare website. Clinicians receive an overall performance score, calculated as a weighted average of performance across these domains, which determines whether they receive positive, negative, or neutral payment adjustments 2 years later.

Performance in the MIPS is assessed primarily from measures selected by clinicians or their practices. For example, practices can report any 6 of nearly 300 measures in the quality domain, which accounts for nearly half of the overall performance score. This flexibility was intended to promote broad participation in the MIPS and stimulate quality improvement in diverse clinical settings. However, allowing practices to select the measures they report raises concerns. First, measure selection undermines a central goal of public reporting: to compare clinicians and practices on a consistent set of measures. Second, because scoring in the MIPS is based on relative performance (i.e., how a practice compares with others on a given measure), practices have a strong incentive to select measures on which they expect to perform well relative to other practices. Thus, measure selection raises concerns about gaming because practices may be able to earn bonuses or avoid penalties by choosing measures on which they already score well, rather than by investing time and resources to provide better care.

Citing these concerns, the Medicare Payment Advisory Commission recommended replacing the MIPS with a program focused on a smaller set of mandatory performance measures, including patient experiences with care. Such a change would effectively make public reporting mandatory for a core set of measures and eliminate practices’ ability to select which measures affect payment adjustments. However, it remains unclear whether practices select measures on which they already score well, and if mandating public reporting on patient experience measures may be beneficial.

To address these questions, we studied precursors of the MIPS—the Value-Based Payment Modifier (VM) and Physician Quality Reporting System (PQRS)—the configurations of which enabled us to examine practice measure selection and mandatory public reporting. We conducted 2 analyses focused on patient experience measures from the Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey. First, we studied measure selection among large practices (≥100 clinicians), which were required to publicly report CAHPS measures starting in 2014 but could voluntarily include these measures in a separate pay-for-performance program. We examined the association between practices’ baseline CAHPS scores and subsequent selection of these measures for quality scoring in the pay-for-performance program. Second, we used the introduction of the CAHPS reporting mandate for large practices and a difference-in-differences design to examine the relationship between mandatory public reporting and changes in patient experiences with care.
Methods

This cross-sectional study was approved by the Harvard Medical School Committee on Human Studies, which granted a waiver of informed consent because we analyzed deidentified secondary data. We followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines for cohort studies. Analyses were conducted between October 1, 2020, and July 30, 2021.

Policy Context

We studied 2 programs that were direct predecessors of the MIPS: the Value-Based Payment Modifier (VM), a pay-for-performance program, and the Physician Quality Reporting System (PQRS), a public reporting program.

The VM was introduced as a voluntary program in 2013 and was fully phased in as a mandatory program for practices with 10 or more clinicians by 2015. Under the VM, practices received upward, downward, or neutral payment adjustments based on their performance in 2 areas: quality of care and per-patient spending. Practices were evaluated on a mandatory set of quality and spending measures but could select additional measures, including CAHPS measures, that contributed to an overall quality score. A prior study examined the association between pay-for-performance incentives in the VM and performance on the program’s mandatory measures. Here, we examined whether practices included CAHPS measures as an optional component of their overall VM quality score.

The PQRS was introduced in 2007 and initially paid bonuses to practices that voluntarily reported performance measures, a subset of which were displayed online on Physician Compare. Over time, components of the PQRS became mandatory. Beginning in 2014, large practices with 100 or more clinicians were required to publicly report CAHPS measures to avoid penalties. Because compliance with this reporting mandate grew rapidly after the first year, we omitted 2014 as a transition year and examined changes in patient experiences from 2011 to 2013 to 2015 to 2016. Additional details about the VM and PQRS are in eTable 1 of the Supplement.

Practices that participated as ACOs in the Medicare Shared Savings Program (MSSP) were not initially included in the VM and reported patient experiences through a separate CAHPS for ACOs survey. Therefore, we excluded practices that participated in the MSSP in any year from 2012 (the program’s first year) through 2016. The MIPS replaced the VM and PQRS in 2017 and made public reporting of patient experiences optional for all practices.

Data Sources and Study Population

Measure Selection and Gaming

We analyzed practice-level data using 2014 to 2016 VM Practice Files, which included the CAHPS scores of large practices that reported these measures for the PQRS. These scores were calculated by CMS based on responses to the CAHPS for PQRS survey, which was administered by a third party to individuals with traditional (ie, fee-for-service) Medicare. Medicare beneficiaries were sampled from practices where they received most primary care. We observed annual practice-level scores for 11 patient experience domains and whether practices elected to include these scores in their overall VM quality score (practices could include all or none of the domain scores).

Practices were informed of their CAHPS scores after these measures could have contributed to the practice’s annual VM quality score. To assess how practice decisions to include CAHPS scores in the VM changed over time as practices learned about their prior performance, we studied a sample of practices that reported CAHPS measures in 2016 (the last year of the VM and PQRS) and started reporting them in 2014 or 2015.
Changes in Patient Experiences Associated With Mandatory Public Reporting

We analyzed patient-level data from the fee-for-service Medicare CAHPS survey, which is separate from but closely related to the CAHPS for PQRS survey (eTable 2 in the Supplement).16 The fee-for-service Medicare CAHPS survey is administered annually to representative samples of community-dwelling Medicare beneficiaries.25,26 Because the survey is conducted early each year and asks respondents to rate their care over the prior 6 months, we analyzed surveys administered from 2012 to 2014 and 2016 to 2017 to assess patient experiences in 2011 to 2013 and 2015 to 2016, respectively, omitting the 2015 survey pertaining to the 2014 transition year.

We used linked Medicare claims from the year before the survey to attribute each respondent to the practice (identified by its taxpayer identification number) that accounted for the plurality of the respondent’s office visits with primary care clinicians in that year.27,28 We excluded 25.7% of respondents without primary care claims and 8.9% of respondents for whom practice size could not be determined. Of remaining respondents, we excluded 36.0% whose practices participated in the MSSP in any year from 2012 to 2016. We limited each annual sample to patients attributed to practices with 50 to 89 or 111 to 150 clinicians, excluding practices with 90 to 110 clinicians to mitigate attenuation bias from small year-to-year fluctuations in practice size that could have changed exposure to the reporting mandate.

Outcome Variables

In analyses of measure selection, our primary outcome was an indicator that a practice elected to include CAHPS scores in its overall VM quality score. In analyses of mandatory public reporting, we examined patient experience measures from 5 domains of the fee-for-service Medicare CAHPS survey that corresponded to domains from the CAHPS for PQRS survey: rating of primary physician, physician communication, timely access to care, access to specialists, and care coordination (eTable 2 in the Supplement). We also analyzed a composite patient experience score, which we calculated as an equally weighted average of scores for the 5 domains.

Practice Size

We measured practice size as the number of clinicians billing under a TIN, consistent with how CMS defined practice size for the VM and PQRS. In analyses of measure selection by large practices, we identified practices with 100 or more clinicians from practice sizes reported in the VM Practice File. In analyses of public reporting, we measured annual practice size using Medicare Provider Practice and Specialty (MD-PPAS) files.29 We measured practice size in the year prior to the fee-for-service Medicare CAHPS survey to align with the period for which we attributed patients to practices.

Respondent Variables

Analyses of the fee-for-service Medicare CAHPS survey included respondents’ demographic characteristics (eg, age, sex, and race and ethnicity), health status, and proxies for socioeconomic status (enrollment in Medicaid and the Medicare Savings Programs30). Race and ethnicity were assessed from the RTI race variable, which classifies Medicare beneficiaries’ race and ethnicity based on Social Security Administration data and an imputation algorithm that identifies additional Hispanic and Asian beneficiaries.31 Variable descriptions are in section 5 in the Supplement.

Statistical Analyses

Measure Selection and Gaming

We categorized large practices that reported CAHPS measures for the PQRS into quintiles of their baseline performance, defined as an equally weighted average of scores on 11 patient experience domains in the first year that practices reported CAHPS measures for the PQRS (2014 or 2015 in our sample). We compared the proportion of practices that included CAHPS scores, measured in the baseline year and up to 2 years later, in their annual VM quality scores, across quintiles of baseline scores. To examine the relationship between baseline scores and future performance, we calculated
correlations of practice-level scores across years. Analyses were conducted using SAS statistical software (version 9.4; SAS Institute, Inc).

Changes in Patient Experiences Associated With Mandatory Public Reporting

We conducted a difference-in-differences analysis to assess changes in patient experiences associated with mandatory public reporting. Specifically, we compared changes in patient experiences from a preintervention period (2011-2013) to a postintervention period (2015-2016) between large practices (111-150 clinicians) affected by the reporting mandate and smaller unaffected practices (50-89 clinicians). For each patient experience score, we estimated a linear difference-in-differences model:

\[ E(\text{Score}_{i,t,k,c,h}) = \beta_0 + \beta_{\text{LargePractice}_k} + \beta_{(2015 \text{ or } 2016)_t} + \beta_{\text{LargePractice}_k \times (2015 \text{ or } 2016)_t} + \beta_{X_{it}} + \beta_{\text{MA}_c,t} + \text{year}_t + \text{HRR}_h \]

where \( \text{Score}_{i,t,k,c,h} \) is a score for respondent \( i \) in year \( t \) who was attributed to practice \( k \) and lived in county \( c \) and Hospital Referral Region \( h \); \( \text{LargePractice}_k \) indicates that practice \( k \) had 111 to 150 clinicians; and \((2015 \text{ or } 2016)_t\) denotes the postintervention period. We adjusted for respondents’ health, demographic, and socioeconomic characteristics \( (X_{it}) \), to account for any compositional changes among patients of large vs smaller practices over time. We adjusted for the annual Medicare Advantage rate by county \( (\text{MA}_c,t) \) to control for potential spillovers of the Medicare Advantage program onto fee-for-service Medicare beneficiaries\(^{32,33} \); year fixed effects \( (\text{year}_t) \) to control for time trends; and hospital referral region fixed effects \( (\text{HRR}_h) \) to control for time-invariant market factors affecting care for patients across large and smaller practices.

Thus, \( \beta_3 \) represents the adjusted within-HRR differential change in patient experiences associated with mandatory public reporting for large practices (pooled across HRRs), through 2 to 3 years after the mandate’s introduction. To facilitate interpretation, we reported estimates of \( \beta_3 \) scaled by the practice-level standard deviations (SDs) of preintervention period scores (termed effect sizes). We adjusted for survey weights and used robust variance estimation to account for clustering within practices. Section 5 in the Supplement provides additional information about these analyses and interpretation of estimates. The threshold for statistical significance was \( P < .05 \) using 2-sided tests.

To examine practice compliance with the reporting mandate, we assessed rates of public reporting of CAHPS measures among large vs smaller practices during 2015 to 2016, which we compared with rates in 2014. We also conducted 2 tests of assumptions underlying the difference-in-differences design.\(^{34} \) First, we estimated differential changes in patient- and practice-level characteristics between large and smaller practices before and after 2014.\(^{35} \) The absence of differential changes supports the assumption that the difference-in-differences model isolates changes in patient experiences associated with public reporting from compositional changes among practices or their patients. Second, we compared preintervention trends in patient experience scores between large and smaller practices. Similar preintervention trends support the assumption that differences in scores between large and smaller practices would have remained constant had the reporting mandate for large practices not been introduced.\(^{34} \)

Sensitivity Analyses

We conducted 2 sensitivity analyses. First, we examined changes in patient experiences from concurrent CAHPS surveys. Second, we reestimated our difference-in-differences models on a broader sample of survey respondents, whom we attributed to practices based on office visits with primary care clinicians or specialists.
Results

Measure Selection and Gaming

Among practices with 100 or more clinicians, 301 publicly reported CAHPS measures in 2016 and started reporting them in 2014 or 2015 (742 practice-years). At baseline (2014 or 2015), these practices had a mean of 431 clinicians and 10,229 patients enrolled in fee-for-service Medicare, among whom the mean age was 71.6 years, 55.8% were women, and 17.8% also received Medicaid (eTable 6 in the Supplement). In 492 (66.3%) of these practice-years, practices voluntarily included CAHPS scores in their overall VM quality score.

Of the 60 practices in the highest quintile of baseline CAHPS scores, 78.9% elected to include these scores in their overall VM quality score for the baseline year, vs 65.9% in the lowest baseline quintile, a difference of 13.0 percentage points (95% CI, 3.9-29.8 percentage points; \(P = .13\); Figure 1). Two years later, 96.3% of practices in the highest baseline quintile included these CAHPS measures in the VM, vs 67.9% of practices in the lowest baseline quintile, a difference of 28.4 percentage points (95% CI, 9.4-47.5 percentage points; \(P = .004\)). In the subset of practices that first reported CAHPS measures in 2014, the 60 practices in the lowest baseline quintile became less likely to include CAHPS measures by 2016, whereas practices in the highest baseline quintile became more likely to include CAHPS measures by 2016 (eFigure 1 in the Supplement). Overall, CAHPS scores were positively correlated within practices across years (eTable 7 in the Supplement).

Changes in Patient Experiences Associated With Mandatory Public Reporting

We analyzed a sample of 21,738 respondents to fee-for-service Medicare CAHPS surveys administered from 2012 to 2014 and 2016 to 2017 (eFigure 2 in the Supplement). In the preintervention period, the mean age of respondents in large practices was 74.1 years, 56.8% of respondents were female, and 6.3% were enrolled in Medicaid (Table 1), closely resembling the community-dwelling Medicare population in the CAHPS sampling frame.26,36 We found few meaningful differential changes in respondent characteristics between large vs smaller practices.
### Table 1. Before Intervention Characteristics and Changes Among Survey Respondents in Large and Smaller Physician Practices

| Respondent characteristics | Before intervention (2011-2013), % | Change from before to after intervention (2011-2013 to 2015-2016) |
|----------------------------|------------------------------------|---------------------------------------------------------------|
|                           | Large practices (111-150 clinicians) [n = 4315 respondents in 344 practices] | Smaller practices (50-89 clinicians) [n = 3999 respondents in 954 practices] |
| Age, y                    | 74.1                               | 74.0                                                         | -0.3 | -0.3 | 0.0 (-0.7 to 0.7) | .99 |
| Sex                       |                                    |                                                              |      |      |                    |    |
| Female                    | 56.8                               | 59.1                                                         | 1.1  | -2.4 | 3.5 (0.6 to 6.5)  | .02 |
| Male                      | 43.2                               | 40.9                                                         | -1.1 | 2.4  | -3.5 (-6.5 to -0.6) | .02 |
| Race and ethnicity        |                                    |                                                              |      |      |                    |    |
| Asian                     | 1.1                                | 0.9                                                          | 0.1  | 0.3  | -0.2 (-1.0 to 0.6) | .62 |
| Black                     | 4.6                                | 5.5                                                          | 2.2  | 0.1  | 2.1 (-2.8 to 7.0) | .39 |
| Hispanic                  | 2.7                                | 2.6                                                          | -0.3 | 0.4  | -0.7 (-2.1 to 0.6) | .28 |
| White                     | 89.3                               | 89.4                                                         | -3.2 | -2.2 | -1.0 (-6.4 to 4.4) | .72 |
| Other                     | 2.4                                | 1.6                                                          | 1.2  | 1.4  | -0.2 (-2.0 to 1.6) | .81 |
| Disabled                  | 16.1                               | 14.7                                                         | -1.7 | 1.8  | 3.5 (-6.2 to -0.8) | .01 |
| End-stage bladder disease | 0.6                                | 0.6                                                          | 0.2  | -0.1 | 0.3 (-0.2 to 0.8) | .18 |
| HCC score                 | 1.3                                | 1.3                                                          | 0.0  | 0.1  | -0.1 (-0.1 to 0.0) | .08 |
| CCW chronic conditions, number | 6.6                                 | 6.6                                                          | 0.1  | 0.1  | 0.0 (-0.3 to 0.2) | .71 |
| Enrolled in Medicaid      | 6.3                                | 6.2                                                          | 1.1  | -0.1 | 1.2 (-1.4 to 3.9) | .36 |
| Enrolled in a Medicare Savings Program | 3.5 | 3.7 | -0.2 | 0.2 | -0.4 (-1.7 to 0.9) | .56 |
| Education                 |                                    |                                                              |      |      |                    |    |
| Less than high school     | 11.5                               | 11.8                                                         | -2.7 | -2.7 | 0.0 (-2.3 to 2.4)  | .98 |
| High school graduate      | 31.1                               | 31.7                                                         | -4.1 | -3.6 | -0.5 (-4.0 to 2.9) | .77 |
| Some college              | 27.2                               | 26.5                                                         | 0.0  | 1.0  | -1.0 (-4.1 to 2.0) | .51 |
| College graduate          | 11.3                               | 10.7                                                         | 2.1  | 1.7  | 0.4 (-2.1 to 2.8)  | .78 |
| Graduate education        | 15.6                               | 15.3                                                         | 3.5  | 3.5  | 0.0 (-2.9 to 3.0)  | .98 |
| Current smoker            | 9.0                                | 9.1                                                          | -1.5 | -0.5 | -1.0 (-2.8 to 0.8) | .28 |
| Use of helper to complete survey | 8.0 | 8.5 | 0.1 | 0.8 | -0.7 (-2.5 to 1.0) | .42 |
| Any functional limitations | 5.2                                | 5.6                                                          | 0.6  | -0.7 | 1.3 (0.0 to 2.6)  | .04 |
| Self-reported general health score | 3.1 | 3.1 | 0.1 | 0.1 | 0.0 (-0.1 to 0.1) | .60 |
| Self-reported mental health score | 3.7 | 3.7 | 0.0 | 0.0 | 0.0 (-0.1 to 0.1) | .97 |

Abbreviation: CCW, Chronic Condition Data Warehouse; HCC, Hierarchical Condition Category.

- Characteristic of respondents to the 2012 to 2014 and 2016 to 2017 fee-for-service Medicare Healthcare Providers and Systems (CAHPS) surveys, as assessed from the survey and Medicare enrollment and claims data from the year prior to the survey (unless otherwise noted). Respondents attributed to practices where they received the majority of their primary care visits in the year prior to the survey. All estimates adjusted for CAHPS survey weights. Respondent characteristics for categorical variables are reported as percentages, which allows us to appropriately display survey-weighted distributions of patient characteristics in the population, rather than numerical frequencies which would not incorporate survey weighting.

- Number of respondents and practices in the preintervention period (2011-2013). In the postintervention period (2015-2016), the sample included 2989 respondents in 262 large practices (111-150 clinicians) and 4999 respondents in 635 smaller practices (50-89 clinicians). Across the preintervention and postintervention periods, the sample included 21738 respondents in 2186 practices.

- That is, changes in the mean or proportion of patients with the characteristic shown in the table row between large and smaller practices from the preintervention to the postintervention periods as defined in the table columns. We estimated these differential changes by fitting a respondent-level linear difference-in-differences model for each characteristic as a function of a postintervention period indicator, an indicator that a patient’s practice had 111 to 150 clinicians, and an interaction between these indicators. Differential changes are given by the regression coefficient on the interaction term. The 95% CIs and P values were calculated using robust standard errors clustered by practice (taxpayer identification number).

- That is, enrollment in full Medicaid.

- Enrollment in 1 of the Medicare Savings Programs, which are partial Medicaid benefits that pay for Medicare Part B premiums, and in some cases Parts A and B cost sharing, for Medicare beneficiaries with low incomes and assets (Section 5 in the Supplement).

- A small proportion (approximately 4%) of respondents did not report their education. We retained these observations in regression analyses by including an indicator variable for missing education status.

- That is, proportion reporting difficulty with 1 or more activities of daily living (bathing, dressing, eating, using chairs, walking, and using the toilet).

- Assessed on a scale of 1 to 5, where 1 indicates poor self-rated health or mental health and 5 indicates excellent self-rated general or mental health.
from the preintervention to postintervention periods (Table 1; eFigure 3 in the Supplement), and no statistically significant differential changes in practice characteristics (eTable 8 in the Supplement).

From 2014 to 2015 to 2016, the proportions of large and smaller practices that reported patient experience measures increased by 36.7 and 4.4 percentage points, respectively, constituting a differential increase among large practices of 32.3 percentage points (95% CI, 23.6-41.0 percentage points; \( P < .001 \); eFigure 4 in the Supplement).

Patient experiences did not differentially improve in large vs smaller practices from the preintervention period to 2 to 3 years after the public reporting mandate (Figure 2). The adjusted estimate for the differential change in the composite patient experience score was \(-0.03\) practice-level SDs of the score (95% CI, \(-0.64\) to 0.58; \( P = .92 \); Table 2). Results were similar in analyses of individual domain scores (Table 2) and in sensitivity analyses (eTables 9 and 10 in the Supplement). Preintervention trends in scores were comparable between large and smaller practices (eFigure 5 in the Supplement).

**Discussion**

In this study of US clinician practices that participated in precursors of the MIPS, we found that large practices were more likely to select CAHPS measures for quality scoring under pay-for-performance when the practices had previously scored well on these measures. However, mandatory public reporting on CAHPS measures was not associated with improved performance on these measures after 2 to 3 years.

The patterns of measure selection we detected were consistent with gaming because a practice would have increased its chances of receiving a bonus or avoiding a penalty by selecting measures on which it expected to perform well, independent of actual quality improvement efforts. For practices, prior CAHPS scores constituted a reliable signal of their future performance, given the high correlation of their scores across years. Accordingly, we found that practices with higher initial CAHPS scores became more likely to include these measures in the pay-for-performance program over time, consistent with strategic measure selection informed by knowledge of prior performance.

![Figure 2. Mean Annual Composite Patient Experience Scores in Large vs Smaller Practices](image-url)
This evidence underscores concerns about gaming in the MIPS, where practices have even greater latitude to select performance measures that affect payment adjustments.

Therefore, the findings of this study support recommendations to end measure selection in the MIPS. Measure selection undermines a central goal of public reporting, which is to facilitate comparisons across practices on the same measures, and is wasteful to the extent that it enables practices to earn bonuses or avoid penalties without improving care. Our findings add to evidence about how practices strategically respond to voluntary components of pay-for-performance programs. For example, in an analysis of the first year of the VM, when practices could voluntarily receive payment adjustments tied to performance, Joynt and colleagues found that practices with better performance scores were more likely to accept performance-based payment incentives.

However, our results also suggest that mandating public reporting of patient experience measures, as recommended in some MIPS reform proposals, may not improve care. Our finding that patient experiences did not improve under mandatory public reporting is consistent with other studies of public reporting programs for clinicians and hospitals, which found little change in quality of care, reflected in process and health outcome measures, under these programs. To our

Table 2. Difference-in-Differences Estimates for Association Between Mandatory Public Reporting and Patient Experiences With Care

| Patient experience scores | Before intervention (2011-2013) | Difference-in-differences estimatesd | P valuef |
|---------------------------|---------------------------------|----------------------------------|---------|
|                           | Mean scores among large practices (111-150 clinicians)b | SDs of practice scoresc | Regression estimate | Effect size, SDs (95% CI)e,f |         |
| Composite scorea          | 80.6                            | 1.6                              | -0.05   | -0.03 (-0.64 to 0.58) | 0.92   |
| Domain-specific scoresb   |                                  |                                  |         |                       |        |
| Rating of primary physician | 90.1                             | 1.0                              | -0.17   | -0.16 (-1.12 to 0.79) | 0.74   |
| Physician communication   | 88.2                             | 1.3                              | 0.19    | 0.15 (-0.68 to 0.97)  | 0.73   |
| Timely access to care     | 66.2                             | 2.3                              | -0.24   | -0.10 (-0.55 to 0.34) | 0.65   |
| Access to specialists     | 84.8                             | 2.5                              | 0.95    | 0.38 (-0.49 to 1.25)  | 0.39   |
| Care coordination         | 80.4                             | 1.5                              | 0.48    | 0.32 (-0.38 to 1.01)  | 0.37   |

Abbreviations: CAHPS, Consumer Assessment of Healthcare Providers and Systems; HRR, Hospital Referral Region; MA, Medicare Advantage.

a Patient experiences with care assessed from the fee-for-service Medicare CAHPS survey. We used responses to surveys administered from 2012 to 2014 and 2016 to 2017 to assess patient experiences with care from 2011 to 2013 and 2015 to 2016, respectively. We omitted the 2015 survey, pertaining to patient experiences in 2014, as a transitional year.

b Mean scores among practices with 111 to 150 clinicians in the preintervention period, adjusted for respondent characteristics in Table 1, annual county-level MA penetration rates, HRR fixed effects, year fixed effects, and survey weights. Scores are standardized to a 0 to 100 scale, with higher scores representing better patient experiences with care (Section 3 in the Supplement).

c Standard deviation (SD) of the practice-level distribution of patient experience scores, estimated among all practices in the preintervention period (Section 5 in the Supplement for details of this calculation).

d Difference-in-differences estimates represent the differential change in composite or domain-specific patient experience scores between large practices (111-150 clinicians) and smaller practices (50-89 clinicians) from the preintervention period (2011-2013) to the postintervention period (2015-2016), adjusted for respondent characteristics in Table 1 and survey weights.

e Effect sizes are difference-in-differences estimates scaled by the practice-level SD of each score. An effect size of -0.16 SDs is equivalent to the difference between the median practice (50th percentile) and a practice at the 44th percentile of performance. The corresponding 95% CIs are also scaled by the practice-level SD in each score.

f 95% CIs and P values were calculated using robust standard errors clustered by practice (taxpayer identification number).

g Calculated at the patient level as an equally weighted average of items comprising all domain-specific scores (Section 3 in the Supplement).

h Calculated at the patient level as an equally weighted average of items within each domain-specific score (Section 3 in the Supplement).
knowledge. No prior studies of public reporting programs examined performance on patient experiences measures. These measures capture aspects of care that are reported directly by patients and may be more comprehensible to patients than technical aspects of care (e.g., process measures), underscoring their importance as quality indicators.44

These findings remain salient amid forthcoming changes to the MIPS. In 2022, CMS will launch the MIPS Value Pathways framework. This framework is intended to reduce reporting burdens by defining core sets of measures, pertaining to specific conditions or specialties, that practices can report. However, MIPS Value Pathways framework may not eliminate gaming, since CMS has indicated that practices will be able to choose which measures (from a given measure set) they report.45

Limitations
Our study had some limitations. First, the extent to which practices strategically selected measures under pay-for-performance was likely diminished by the uncertain payoff (i.e., increase in bonuses or reduction in penalties) associated with selecting specific measures, since in the VM payment adjustments were based on a composite of performance measures assessed among all practices, and the set of measures and practices varied year-to-year.46 However, practices in the MIPS face similar uncertainty about the set of peers reporting on a given measure, and thus are also likely to rely on assessments of their own past performance relative to other practices when selecting measures.46

Second, findings from our difference-in-differences analyses may not generalize to Medicare beneficiaries living in institutional settings, such as nursing homes, because the fee-for-service Medicare CAHPS survey samples from the community-dwelling Medicare population. Third, our difference-in-differences analyses could have been biased by unmeasured confounders. However, we did not detect meaningful differential changes on observable patient or practice characteristics.

Conclusions
In precursors of the MIPS, we found that large clinician practices were more likely to voluntarily include patient experience measures in a pay-for-performance program when they previously performed well on these measures. However, mandatory reporting of patient experiences was not associated with improved performance on these measures. These findings underscore concerns about gaming in the MIPS and provide cautionary evidence about proposed reforms to this program.
Drafting of the manuscript: Roberts, Song.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Roberts, Ding, McWilliams.

Obtained funding: Song, McWilliams.

Supervision: McWilliams.

Conflict of Interest Disclosures: Dr Roberts reported grants from Agency for Healthcare Research and Quality K01HS026727 and grants from Arnold Ventures during the conduct of the study; grants from Arnold Ventures (separate from grant acknowledged under 2A) outside the submitted work. Dr Song reported grants from National Institutes of Health, Office of the Director NIH Director’s Early Independence Award, DP5-OD0024564 and grants from Laura and John Arnold Foundation during the conduct of the study; grants from National Institute on Aging outside the submitted work; and personal fees from the Research Triangle Institute for work on Medicare risk adjustment, from Google Ventures and the International Foundation of Employee Benefit Plans for academic lectures outside of this work; and for providing consultation in legal cases. Dr McWilliams reported grants from Arnold Ventures and grants from National Institute on Aging during the conduct of the study. No other disclosures were reported.

Funding/Support: Supported by grants from Arnold Ventures, the National Institutes of Health (PO1 AG032952), and the Agency for Healthcare Research and Quality (K01 HS026727).

Role of the Funder/Sponsor: Arnold Ventures, the National Institutes of Health, and the Agency for Healthcare Research and Quality had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclaimer: This work does not necessarily reflect the views of the National Institutes of Health, Arnold Ventures, or the Agency for Healthcare Research and Quality.

REFERENCES
1. Schneider EC, Hall CJ. Improve quality, control spending, maintain access—can the merit-based incentive payment system deliver? N Engl J Med. 2017;376(8):708-710. doi:10.1056/NEJMp1613876
2. 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(214):77008-77831.
3. Chapter 2: Medicare’s new framework for paying clinicians. In: Report to the Congress: Medicare and the Health Care Delivery System. Washington DC: Medicare Payment Advisory Commission; 2016:29-54.
4. 2018 Quality Payment Program Experience Report. Baltimore, MD: Centers for Medicare and Medicaid Services; October 28, 2020. Accessed October 28, 2021. https://data.cms.gov/quality-of-care/quality-payment-program-experience
5. Medicare Program; CY 2020 Revisions to Payment Policies Under the Physician Fee Schedule and Other Changes to Part B Payment Policies. Services CMM. Vol 84 FR 62568. Washington, DC2020-62568-63563. Accessed August 19, 2020. www.federalregister.gov/d/2019-24086
6. Quality Payment Program: Quality Measures. Centers for Medicare and Medicaid Services. Accessed August 19, 2020. https://qpp.cms.gov/mips/explore-measures?tab=qualityMeasures&py=2020#measures
7. 2019 Merit-based Incentive Payment System (MIPS) Quality Performance Category Fact Sheet. Baltimore, MD: Centers for Medicare and Medicaid Services; July 13, 2020. Accessed August 19, 2020. https://qpp-cm-prod-content.s3.amazonaws.com/uploads/350/2019%20MIPS%20Quality%20Performance%20Category%20FactSheet.pdf
8. Chapter 15: Moving beyond the Merit-based Incentive Payment System. Washington DC: Medicare Payment Advisory Commission; 2018.
9. Rathi VK, McWilliams JM. First-year report cards from the Merit-Based Incentive Payment System (MIPS): what will be learned and what next? JAMA. 2019;321(12):1157-1158. doi:10.1001/jama.2019.1295
10. Merit-Based Incentive Payment System: Scoring 101 Guide for the 2019 Performance Year. Baltimore, MD: Centers for Medicare and Medicaid Services;2020.
11. McWilliams JM. Macra: big fix or big problem? Ann Intern Med. 2017;167(2):122-124. doi:10.7326/M17-0230
12. Medicare Shared Savings Program Interaction with the 2017 Value Modifier: Frequently Asked Questions. Baltimore, MD: Centers for Medicare and Medicaid Services; September 2016 2016. Accessed June 3, 2021. https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/2017-VMM-SSP-FAQs.pdf
13. Rosenthal MB, Frank RG. What is the empirical basis for paying for quality in health care? Med Care Res Rev. 2006;63(2):135-157. doi:10.1177/1077558705285291

14. Roberts ET, Zaslavsky AM, McWilliams JM. The value-based payment modifier: Program outcomes and implications for disparities. Ann Intern Med. 2018;168(4):255-265. doi:10.7326/M17-1740

15. Koltov MK, Damle NS. Health policy basics: physician quality reporting system. Ann Intern Med. 2014;161(5):365-367. doi:10.7326/M14-0786

16. CAHPS for Physician Quality Reporting System (PQRS) Survey. Accessed September 9, 2021. www.pqrscahps.org/en/about-the-survey/

17. CAHPS for Physician Quality Reporting System (PQRS) Survey. Accessed September 9, 2021. www.pqrscahps.org/en/about-the-survey/

18. Moen EL, Bynum JPW. Evaluation of physician network-based measures of care coordination using Medicare patient-reported experience measures. J Gen Intern Med. 2019;34(11):2482-2489. doi:10.1007/s11606-019-05313-y

19. 2016 Value-Based Payment Modifier Program Experience Report. Baltimore, MD: Centers for Medicare and Medicaid Services; April 2017. Accessed August 20, 2021. www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/2016-VMP-Experience-Report.pdf

20. 2015 Reporting Experience Including Trends (2007-2015): Physician Quality Reporting System. Baltimore, MD: Centers for Medicare and Medicaid Services; 2017. Accessed August 20, 2021. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/Downloads/2015_PQRS_Experience_Report.pdf

21. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Equator Network. Accessed July 30, 2021. https://www.equator-network.org/contact/contact/

22. 2020 CAHPS for MIPS Survey via CMS-Approved Survey Vendor Reporting. Centers for Medicare and Medicaid Services. Accessed August 19, 2020. https://qpp-cm-prod-content.s3.amazonaws.com/uploads/925/2020%20CAHPS%20for%20MIPS%20Overview%20Fact%20Sheet.pdf

23. Mugge A. Physician Quality Reporting System (PQRS). Medicare Learning Network Connects. Published Undated. Accessed March 31, 2020. https://fdocuments.in/document/physician-quality-reporting-system-pqrs-electronic-clinical-quality-questions.html

24. 2016 Physician Quality Reporting System (PQRS): CMS-Certified Survey Vendor Reporting Consumer Assessment of Healthcare Providers and Systems (CAHPS) for PQRS Made Simple. Accessed September 9, 2021. www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/downloads/2016PQRS_CAHPS_MadeSimple.pdf

25. Fee-for-Service (FFS) CAHPS. Centers for Medicare and Medicaid Services. Access. https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/CAHPS/FFSCAHP5

26. Medicare Advantage, Medicare Part D, and Medicare Fee-For-Service Consumer Assessment of Healthcare Providers and Systems (CAHPS) Survey: Supporting Statement Part B. Baltimore, MD: Centers for Medicare and Medicaid Services; February 2, 2021 2021. Accessed August 20, 2021. www.cms.gov/Regulations-and-Guidance/Legislation/PaperworkReductionActof1995/PRA-Listing-Items/CMS-R-246

27. McWilliams JM, Landon BE, Chernew ME, Zaslavsky AM. Changes in patients’ experiences in Medicare accountable care organizations. N Engl J Med. 2014;371(18):1715-1724. doi:10.1056/NEJMsa1406552

28. Roberts ET, Mehrotra A, McWilliams JM. High-price and low-price physician practices do not differ significantly on care quality or efficiency. Health Aff (Millwood). 2017;36(5):855-864. doi:10.1377/hlthaff.2016.1266

29. Medicare Data on Provider Practice and Specialty (MD-PPAS) User Documentation, Version 2.2. Baltimore, MD: Centers for Medicare and Medicaid Services; February 2017 2017. Accessed July 30, 2021. https://resdac.org/cms-data/variables/research-triangle-institute-rti-race-code

30. Samson LW, Finegold K, Ahmed A, Jensen M, Filice CE, Joynt KE. Examining measures of income and poverty in medicare administrative data. Med Care. 2017;55(12):e158-e163. doi:10.1097/MLR.00000000000000606

31. Research Triangle Institute (RTI) Race Code. Research Data Assistance Center. Accessed July 30, 2021. https://resdac.org/cms-data/variables/research-triangle-institute-rti-race-code

32. Glied S, Zivin JG. How do doctors behave when some (but not all) of their patients are in managed care? J Health Econ. 2002;21(2):337-353. doi:10.1016/S0167-6296(03)00313-X

33. Baicker K, Chernew ME, Robbins JA. The spillover effects of Medicare managed care: Medicare Advantage and hospital utilization. J Health Econ. 2013;32(6):1289-1300. doi:10.1016/j.jhealeco.2013.09.005
34. Daw JR, Hatfield LA. Matching and regression to the mean in difference-in-differences analysis. Health Serv Res. 2018;53(6):4138-4156. doi:10.1111/1475-6773.12993

35. McWILLIAMS JM, Hatfield LA, Landon BE, Chernew ME. Savings or selection? initial spending reductions in the Medicare Shared Savings Program and considerations for reform. Milbank Q. 2020;98(3):847-907. doi:10.1111/1468-0013.00348

36. 2018 Medicare Current Beneficiary Survey Public Use File. In: Services CMM. Baltimore, MD2020. Accessed August 20,2021. https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS

37. Berdahl CT, Easterlin MC, Ryan G, Needleman J, Nuckols TK. Primary Care Physicians in the Merit-Based Incentive Payment System (MIPS): a Qualitative Investigation of Participants’ Experiences, Self-Reported Practice Changes, and Suggestions for Program Administrators. J Gen Intern Med. 2019;34(10):2275-2281. doi:10.1007/s11606-019-05207-z

38. Joynt Maddox KE, Epstein AM, Samson LW, Chen LM. Performance And Participation Of Physicians In Year One Of Medicare’s Value-Based Payment Modifier Program. Health Aff (Millwood). 2017;36(12):2175-2184. doi:10.1377/hlthaff.2017.0894

39. Ryan AM, Nallamothu BK, Dimick JB. Medicare’s public reporting initiative on hospital quality had modest or no impact on mortality from three key conditions. Health Aff (Millwood). 2012;31(3):585-592. doi:10.1377/hlthaff.2011.0719

40. Dowd BE, Swenson T, Parashuram S, Coulam R, Kane R. PQRS participation, inappropriate utilization of health care services, and Medicare expenditures. Med Care Res Rev. 2016;73(3):106-123. doi:10.1177/1077558715597846

41. Ryan AM. Effects of the Premier Hospital Quality Incentive Demonstration on Medicare patient mortality and cost. Health Serv Res. 2009;44(3):821-842. doi:10.1111/j.1475-6773.2009.00956.x

42. Dragone D, Kessler D, McClellan M, Satterthwaite M. Is more information better? the effects of "report cards" on health care providers. J Polit Econ. 2003;111(3):555-588. doi:10.1086/374180

43. Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. JAMA. 2005;293(10):1239-1244. doi:10.1001/jama.293.10.1239

44. Anhang Price R, Elliott MN, Zaslavsky AM, et al. Examining the role of patient experience surveys in measuring health care quality. Med Care Res Rev. 2014;71(5):522-554. doi:10.1177/1077558714541480

45. Medicare Program; CY 2022 payment policies under the physician fee schedule and other changes to part b payment policies; Medicare Shared Savings Program requirements; provider enrollment regulation updates; provider and supplier prepayment and post-payment medical review requirements in Washington, DC. Federal Register. 2021:2432-2443.

46. Navathe AS, Dinh CT, Chen A, Liao JM. Findings And Implications From MIPS Year 1 Performance Data. Health Affairs Blog. Vol 20212019. Accessed August 20, 2021. www.ajmc.com/view/physician-perspectives-how-the-merit-based-incentive-payment-system-improves-value

SUPPLEMENT.
Section 1. Policy context

eTable 1. Structure and phase-in of the VM and PQRS by year and clinician practice size

Section 2. Fee-for-service Medicare CAHPS and CAHPS for PQRS surveys

eTable 2. Items and domains in the CAHPS for PQRS survey and corresponding items in the Fee-for-service Medicare CAHPS survey

eTable 3. Correlations of patient experience scores across domains in the CAHPS for PQRS survey

eTable 4. Items with missing responses in the Fee-for-service Medicare CAHPS survey

Section 3. Analysis of patient experience scores in the Fee-for-service Medicare CAHPS survey and concordance with practice scores from the CAHPS for PQRS survey

Construction of domain and composite patient experience scores in the Fee-for-service Medicare CAHPS survey

Concordance of patient experiences in the Fee-for-service Medicare CAHPS survey with practice-level scores from the CAHPS for PQRS survey

eTable 5. Composite patient scores from the Fee-for-service Medicare CAHPS survey by quintile of practice scores

Section 4. Measure selection and gaming analysis

eTable 6. Characteristics of large practices and their patients, among large practices that reported CAHPS scores in 2016 and began reporting them in 2014 or 2015

eTable 7. Practice-level correlations of CAHPS scores across years

Section 5. Difference-in-differences analysis

Respondent sample for difference-in-differences analyses

eFigure 1. Proportions of practices including CAHPS scores in the VM in 2014-2016 (N=140 practices that first reported CAHPS scores in 2014)
eFigure 2. Sample inclusion criteria for difference-in-differences analysis
Respondent characteristics
eFigure 3. Means or proportions of patient characteristics in large vs. smaller practices from 2011-2013 and 2015-2016
eFigure 4. Proportions of large and smaller practices publicly reporting CAHPS measures in 2014 vs. 2015-2016
eFigure 5. Event-study plots of annual differential changes in composite and domain-level patient experience scores between large and smaller practices (relative to 2013)
eTable 8. Preintervention practice characteristics and changes in the characteristics of large vs. smaller practices in difference-in-differences analysis
eTable 9. Difference-in-differences estimates based on responses to concurrent Fee-for-service Medicare CAHPS surveys
eTable 10. Difference-in-differences estimates among patients attributed to practices based on outpatient claims with primary care clinicians or specialists