MEANINGFUL USE: CLINICIANS' QUALITY SCORES & CONSUMER RATINGS IN THE MERIT-BASED INCENTIVE PAYMENT SYSTEM

AN HONORS THESIS

SUBMITTED ON THE 27TH DAY OF APRIL, 2020

TO THE DEPARTMENT OF ECONOMICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

OF THE HONORS PROGRAM

OF NEWCOMB-TULANE COLLEGE

TULANE UNIVERSITY

FOR THE DEGREE OF

BACHELOR OF ARTS

WITH HONORS IN ECONOMICS

BY

Emily Jeanne Hart

APPROVED:

ENGY ZIEDAN

Engy Ziedan, Ph.D.
Director of Thesis

LaPorchia A. Collins, Ph.D.
Second Reader

Mollye M. Demosthenidy, J.D., M.H.A.
Third Reader

Abstract

In 2017, the Center for Medicare and Medicaid Services (CMS) implemented its Quality Payment Program as a means to reimburse clinicians on the basis of clinical quality rather than solely on the quantity of provided health services. This policy replaces the previous payment models with a value-based purchasing model to incentivize more patient-centered and efficient healthcare amongst clinicians. Defining and measuring health provider "quality" remains ambiguous. My research empirically examines the link between patient-reported quality measures and technical quality measures in the program. Specifically, I use a regression analysis with fixed effects to study the two quality ratings from the Merit-based Incentive Payment System (MIPS) in outpatient settings. I find that the correlation between the quality rating scores is statistically significant ($p < 0.001$), but the association is weak (the technical quality score accounts for 0.004% of the total magnitude of the patient-rating score). Existing literature supports these findings and concludes that each score plays a unique role in capturing the clinician's "true" quality. Findings of this research could inform whether or not patients value care apart from what is clinically appropriate.

Acknowledgments

Completing a Senior Honors Thesis in the field of Healthcare Economics involves conquering such a steep learning curve that I could not dream of doing this alone. What follows is my heartfelt thank you to those who willingly and, perhaps at times, unwillingly, but always compassionately contributed to the creation of this thesis. I first wish to express my sincere appreciation to Dr. Engy Ziedan, my thesis director, whose decisive and smart input and kind reminders to not let "great' stand in the way of 'good" helped me to persevere through the many hours of research and analysis into the multifaceted area of quality reporting. I feel privileged to have been able to work closely with someone of her expertise in health economics and payment reform. I owe a debt of gratitude to my second reader, Dr. LaPorchia Collins for her clear and meaningful editorial suggestions to my thesis. Her attentiveness throughout this year and kindness as I visited numerous office hours helped bolster my resolve to press on in this thesis journey. Thank you to my third reader, Dr. Mollye Demosthenidy, for giving me a greater understanding and perspective of the dialogue surrounding healthcare reform and novel payment models and to Dr. Michelle Lacey for her pivotal input related to CMS data file statistics. I would be remiss if I did not thank my family and friends for the many late nights of proofreading and pep talks. I am forever grateful to my honorary fourth reader, Sophie Drew, whose revisions and supportive advice throughout this process were truly indispensable. A special thanks to Dr. Emmy Hall Ganos of the Robert Wood Johnson Foundation, whose expert insight about the research topic and encouragement carried me to the finish line. I would lastly like to thank my mom and God, for obvious reasons. Without His providence, this achievement would not have been possible.

Table of Contents

List of Tables

List of Figures

1. Specific Aims

Healthcare systems in many developed countries are restructuring payment models to reward quality, rather than quantity, of provided health services. The Merit-based Incentive Payment System (MIPS) is a value-based purchasing option for Medicare providers that links reimbursement to providers' performance score. MIPS affects nearly 90% of all US healthcare providers (Boccuti et al., 2015). The idea behind MIPS is that physician reimbursement should be a function of quality—primarily, clinical quality. However, policymakers and physicians have echoed concerns that clinical quality alone does not take into account patient evaluations of a medical professional (Mehta, 2015). This research examines the link between patient-reported quality measures and clinical quality measures. I use a case study of the Merit-based Incentive Payment System in outpatient settings to examine this link. Findings of this research could inform whether or not patients value care apart from what is clinically appropriate.

Specifically, I use a series of datasets from the Centers for Medicare and Medicaid Services (CMS) that contain information on the demographics, utilization, and performance ratings of MIPS eligible clinicians and clinician groups. I use this information to estimate the association between patients' self-reported rating of a provider and the provider's final MIPS score. The final MIPS score is a weighted average of the provider's performance on the following four categories: quality, improvement activities, advancing care information, and healthcare costs (CMS, n.d.). For this research, I execute an empirical analysis (regression model) of patients' rating of a provider versus the clinical quality rating of a provider, controlling for provider specialty, provider’s clinical experience in years, mean patient risk score, provider gender, and geographic state.

Conceptually, there are several reasons why clinical quality may not match the patient's perceived quality of a physician. First, health economics theory suggests that patients with low out-of-pocket payments may demand more care than those with higher out-of-pocket payments. Demand increases as the price paid for care decreases. This phenomenon—widely known as moral hazard—suggests that insured patients with low cost-sharing tend to have higher demand for care than those with high cost-sharing or no insurance at all (Manning et al., 1987). The insured individuals may report lower satisfaction and perceived quality scores if the doctor does not or cannot meet their high demands for care. Second, MIPS includes measures that require physicians to report quality data on clinical practice behaviors (CMS, n.d.). Some measures aim to reduce inappropriate or suboptimal care processes while others aim to promote underutilized care processes. These measures could be at odds with patient satisfaction. As a result, doctors who do not meet higher demands of patients in order to align with the MIPS measurement of quality may receive lower patient-perceived quality scores. Conversely, doctors who aim to please patients and score highly on patient-perceived quality may receive low MIPS scores due to their low adherence to care processes.

2. Background

2.1 Value-based Purchasing

Value-based purchasing models replace purely volume-based “fee-for-service” payments to healthcare service providers. In the US, value-based purchasing models have different subtypes, including bundled payments, pay-for-performance, and shared savings (ACP, n.d.; Chee et al., 2016; James, 2012). The Centers for Medicare and Medicaid Services (CMS) introduced value-based purchasing programs nationally via Medicare reimbursements.

The transition to value-based purchasing models represents the Department of Health and Human Services' effort to optimize healthcare quality and cost efficiency (Burwell, 2015). By 2017, about 90% of Medicare's traditional fee-for-service payments and 34% of all healthcare payments were linked to performance of quality measures (Health Care Payment Learning & Action Network, 2018). Other payment models that target excess healthcare cost, such as prospective payments and capitation, do not always consider healthcare quality or outcomes. Value-based purchasing models factor in both cost and utility of healthcare services (Vuong et al., 2018).

Evaluating healthcare spending and payment mechanisms is of particular importance to US policymakers. Voters ranked healthcare reform as the priority issue for the past two presidential elections (The Henry J. Kaiser Family Foundation, 2020). Additionally, in 2018, nearly 18% of the United States' gross domestic product went toward healthcare spending (CMS, n.d.; Hartman et al., 2019). Meanwhile, other OECD countries spend about half as much on healthcare and have better life expectancy, lower number of hospitalizations from preventable causes, and lower rates of avoidable deaths (Tikkanen, Roosa & Abrams, Melinda K, 2020). Restructuring payments to health providers is one of many channels the US government uses to control the cost growth in the healthcare sector.

2.2 Merit-based Incentive Payment System (MIPS)

The Medicare Access and CHIP Reauthorization Act of 2015 (MACRA) established a quality incentive program, termed the Quality Payment Program (QPP) (CMS, 2016). This program is a culmination and combination of other quality-based payment programs of the past, including the Value Modifier and Physician Quality Reporting System (CMS, 2019). The Quality Payment Program offers clinicians a choice between two reimbursement tracks, the Alternative Payment Model (APM) and the Merit-based Incentive Payment System (MIPS). Providers who take Medicare patients must choose between one of the two tracks. MIPS sets *a priori* baseline fee-for-service reimbursement rates and then positively or negatively adjusts this rate for individual physicians. The direction of the payment adjustment depends on the individual physician's final performance score (LAN, 2018). For example, a physician who earns a final MIPS score that is higher than the national average will receive a bonus payment on top of the regular fee-for-service payment. By 2022, a physician who scores below average will receive up to nine percent less than the baseline reimbursement rate for his services (CMS, n.d.).

The final score is a composite score, calculated from the provider's scores in four categories: Quality, Improvement Activities, Advancing Care Information, and Healthcare Costs. The category scores are based on the provider's performance and engagement in select activities or "measures". Providers choose which measures to report and subsequently be scored upon. Before becoming available to physicians, the measures are first endorsed by either a physician specialty society or the National Committee for Quality Assurance (NCQA), which is an independent nongovernmental organization composed of physicians and public health experts. MIPS measures track either a process or patient outcome of care (AAFP, 2019; CMS, n.d).

MIPS process measures require providers to report the completion of certain tasks in every patient visit, for example, taking the patient's blood pressure, reviewing the patient's list of medications, or prescribing antibiotics only after confirmation of a bacterial infection (CMS, n.d.). Other MIPS measures are outcome-oriented and highlight the program's goals of rewarding patient-centered care and patient health outcomes in and of themselves. One such measure is the Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey, which asks patients for feedback about their health visits. In the survey, patients respond to questions about the respectfulness and responsiveness of the office staff, how well the provider communicated, and the overall rating of their provider, among other questions (AHRQ, 2018). The consumer survey is set apart from other patient review platforms because patients are systematically surveyed, and the questions are validated research instruments (Fowler, 1997; Harris-Kojetin et al., 1999; Scholle et al., 2012).

2.3 Spillover Effects

Medicare's Quality Payment Program will likely trigger a response in the private insurance sector to reimburse physicians on care value. In the past, private health insurance plans have followed the prices and standards set by Medicare (Clemens & Gottlieb, 2017; Clemens et al., 2017). As value-based purchasing models become more popular in the health insurance market, the shift in payment and potential increase in administrative burden may raise overhead costs to a level that only large healthcare organizations can sustain due to their economies of scale (Bresnahan & Levin, 2012). Physicians in small private practices may feel mounting pressure to choose between leaving the industry or consolidating with a hospital if their private practice cannot sustain the transition to the new system (Ancker et al., 2013; Ehrlich et al., 2017; Rao, S. K. et al., 2017; Woolhandler & Himmelstein, 2014). Pressuring doctors to make this choice may exacerbate the predicted provider shortage in the US (Zgierska et al., 2014; Zhang et al., 2020). However, the process of implementing a new system creates an industry of healthcare workers that specialize in quality reporting and documentation, which may improve job security and lower the unemployment rate (Martiniano & Moore, February 2018).

Spillover effects are important to consider in light of the broader goal of policy to maximize social welfare. To improve the welfare of both patients and physicians through policy reform, it is also important to contextualize “high quality” healthcare (Hanefeld et al., 2017); hence, the overall aim of my thesis is to examine the relationship between the MIPS clinical quality score and the CAHPS patient-perceived quality score. However, this study does not examine whether the MIPS clinical score is an accurate measure of technical quality of a provider or whether providers are cherry-picking quality indicators that are relatively less costly to improve upon or report than other indicators. Additionally, this study does not address how the CAHPS survey is administered or how the response format, such as using global scales or dichotomous scales, influences patients’ reporting of their healthcare experiences (Drake et al., 2014; Martino et al., 2017). For example, I do not question whether a patient completed the survey because of some perceived pressure or incentive, resulting in skewed ratings of a provider

(Hargraves et al., 2019; Klein et al., 2011). Answers to such questions are a source of future study or have been reviewed by previous literature.

3. Conceptual Framework

Conceptually, physicians who receive high patient-perceived quality scores may also perform well on MIPS clinical quality measures. A physician who is well-trained by her medical school could have a positive correlation in the two scores because physician's education should encompass training in both bedside manner and clinical quality. Other correlates such as age or specialty could drive the positive correlation between the two scores (Chen et al., 2017). It is also scientifically plausible that physicians who are highly ranked by patients actually perform poorly on MIPS clinical quality measures. Theoretically, there are three explanations for a possible discrepancy among patient-perceived quality of providers and MIPS clinical quality of providers: time-inconsistent preferences, patient bias, and physician bias.

For one, time-inconsistent preferences could explain why patients value aspects of care that may be in conflict with what is considered high clinical quality care. When making treatment decisions on behalf of the patient, a physician has a long run objective function in mind and understands what is best for the patient in terms of clinical quality. Meanwhile, the patient has a different objective function in mind that prefers short-run utility and places more value on immediate relief. Some patients may not consider or be aware of potential long-term consequences of their treatment demands. Others could understand the health risks associated with their lifestyle and decide that the cost of improving their chances of living longer are too high. Because they are not willing to make the investment in health that could have a high return in future ages, patients demand care that, while less sustainable, offers a higher return in the current period (Becker, 2007; Oster, 2012). For example, a physician would be concerned about preventing antibiotic resistance or drug addiction in the future while the patient values what they think will give them the fastest relief, whatever amount of treatment that may be.

Secondly, health care consumers may have cognitive biases that favor visible or tangible products in exchange for their copayments or out-of-pocket health care expenses over intangible products (Levitt, 1981). Over the typical exchange of medical advice, patients may have an expectation to receive a prescription, ointment or new device at each clinician visit. In this scenario, patients understand the different utilities of healthcare and specifically prefer the more invasive or intensive treatment option. For example, a patient with high cholesterol who pays a co-pay to visit a doctor may opt for the statin prescription rather than professional advice on lifestyle changes. The more invasive treatment option demanded need not always directly aid in recovery and may even be a placebo treatment (Espay et al., 2015). A patient who does not obtain the visible or intensive treatment that they desire may negatively rate the healthcare experience, even when such care is not clinically appropriate.

Thirdly, the method used by CMS to adjust providers' MIPS scores based on the patient population health risks may not fully reflect the patient's actual health or comorbidities (Iezzoni, 1994; Kumar et al., 2016), potentially leading physicians to shortchange or unduly restrict clinically necessary care to patients who do not appear to have a documented record of high risk in order to satisfy the MIPS clinical score criteria.

Hierarchical Condition Category risk scores account for comorbidities but do not account for other variables that are determinants of health such as socioeconomic status, race, or education (Braveman & Gottlieb, 2014; Quigley et al., 2018).

4. Data and Research Approach

I used three independent datasets that consist of the following: the Physician Compare National Downloadable File, the Physician Compare 2017 Individual EC Public Reporting - Overall MIPS Performance, and the Physician Compare 2017 Group Public Reporting - Patient Experience.

The Physician Compare National Downloadable File contains demographic information about individual clinicians and has 2,220,119 observations in total. I extract 9 of the 38 variables related to the provider including the National Provider Identification (NPI) code, the gender of the provider, the type of medical credential and graduation year, the primary specialty, the legal name of the practice group, and the number of hospital affiliations.

The Physician Compare 2017 Group Public Reporting - Patient Experience file contains the group-level results for the Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey and includes a total of 1,903 observations. The results are listed by the clinician-organization group legal name. I extract 6 of 8 variables, including the legal name of the practice group, the consumer’s state of residence, the measure code, the measure title, the measure performance rate, and the denominator count.

I merge the CAHPS data to the Physician Compare National Downloadable File using the “group legal name” variable. Hence, every clinician who belongs to a particular clinician group is assigned the corresponding group-level scores for each CAHPS category. All but 176 observations in the CAHPS data file match with at least one provider from the Physician Compare National Downloadable File. This combined dataset includes 14 variables and 214,130 observations.

The Physician Compare 2017 Individual EC Public Reporting - Overall MIPS Performance File which contains the Merit-Based Incentive Payment System (MIPS) final composite score and the category performance scores for the individual clinicians results in a total of 413,233 observations. These observations are listed by National Provider Identification (NPI) code. From these observations I extract 2 of the 9 variables which includes the NPI code and the final MIPS score.

I merge the MIPS data to the existing combined dataset using the “NPI” variable. A single observation is represented by an individual clinician with both MIPS and CAHPS scores. Some clinicians may appear more than once in the data because the CAHPS measures vary, and there are multiple measures for each clinician assigned to a group. For those with multiple measures I create a variable of the average score across all of the CAHPS measures for each group, and I isolate the CAHPS measure that reports the patients’ rating of their provider in the group. The dataset used for the empirical analysis has 17 variables and 26,841 observations in total. The observations represent a unique clinician, listed by NPI code. Ultimately, CAHPS data properly merged with 5.31% of the clinicians from the overall MIPS dataset. This rate is sufficient for a robust sample.

The variable of interest is the mean CAHPS score, which is defined as the percent of patients rating the provider a 9 or 10 out of 10 on a global rating scale, where 10 is the highest quality experience. I remove from the sample the lowermost and uppermost values of the CAHPS score (score of 0 and 100, respectively) to prevent the ceiling effect. The ceiling effect in statistics describes an upper limit of scores at which point the variance can no longer be detected (Cramer & Howitt, 2005; Everitt, 2002; Terwee et al., 2007). I assign group CAHPS scores to all the clinicians within the group, assuming that the clinicians associate because they are of like quality (Kash & Tan, 2016). In the sample, 1.1% of clinicians were affiliated with more than one organization; in these cases, the average CAHPS of the two organizations was assigned to the clinician.

Other variables on a per-provider level include the mean patient age, the proportion of male patients, the mean patient risk score, the provider's clinical experience in years, the provider's gender, the provider's specialty, and a dummy variable measuring whether the provider has one or more than one hospital affiliation.

5. Empirical Model

$$\text{CAHPS}_i = \alpha + \beta_1 MIPS_i + \beta_2 Patient\ Avg.\ Risk\ Score_i + \beta_3 MD\ Experience_i + \beta_4 MD\ Gender_i + \beta_5 Specialty_i + \beta_6 State_i + \varepsilon_i$$

i = 1,…, 26841 (provider)
year = 2017

The empirical model uses a fixed-effects Ordinary Least Squares (OLS) regression framework, where Y_i is the mean patient rating CAHPS score of a healthcare provider, *i*. Included on the right-hand side of the equation is: the constant α; controls for mean patient risk score, the provider's clinical experience in years, a provider gender dummy that equals 1 if the doctor is male, a provider specialty fixed effect, and a state fixed effect; and an error term ε_i. The main coefficient of interest is β_1, the correlation of the clinician's MIPS score to her CAHPS score. If $\beta > 0$, an increase in the MIPS score is associated with an increase in the mean patient rating of the provider, on average, holding other variables constant.

For all regression analyses, I begin with a "Base" model that checks the coarse association between the MIPS and CAHPS scores and the fit of the model without controlling for any other factors. I then run two "Moderate" models. The first moderate specification includes only the two controls of average patient risk score and provider experience in order to reveal statistical significance of the two controls without any dummy variable fixed effects. The second moderate specification drops the two continuous variable controls and adds the state, provider gender, and provider specialty fixed effects. The presence of the three fixed effect variables should result in a model that is better fit to the data. Lastly, I evaluate the "Intense" model, which is the most rigorous or parameterized regression, to reduce omitted variable bias. It includes state, provider specialty, and provider gender fixed effects in addition to the set of continuous variables, average patient risk score and provider experience. The equation above represents the "Intense" Model. The results are shown in Table 2. All parameters in this model are included to account for possible factors that may bias the CAHPS score or damage the fit of the model to the data.

State fixed effects are a powerful econometric tool used to find the variation in each state in the year 2017. Namely, I consider state fixed effects to account for the average socioeconomic status of the state population, state healthcare policies, such as if

Medicaid has been expanded, and any statewide sentiment about physician practice behaviors held by patients, such as preferences for less intrusive medical care. The provider gender dummy is included to capture any residual bias in the interaction between patient and doctor that may affect the CAHPS score beyond that of the doctor's clinical quality. The provider specialty dummy is meant to account for variation in scores affected by the fact that some specialties have more patient contact than others. For example, a nurse or internal medicine doctor typically communicates with a patient for a longer period of time in the visit, increasing the probability that the CAHPS survey is completed by patients who are loyal and with whom the doctor has built rapport.

I also include two continuous control variables in the "Moderate Controls Only" panel and "Intense Controls & FE" regression equations. One is the average patient risk score, which accounts for patients who may project their health status onto the performance of the clinician through lower CAHPS scores, as evidenced by previous literature (Hays et al., 2018; Poot et al., 2019). The second control variable is the doctor's years of clinical experience, which captures the sentiment of patients toward more experienced clinicians or toward clinicians to whom the patients have been loyal, all of which could falsely increase satisfaction scores of more experienced clinicians (Peck, 2011).

I introduce other variables to check the sensitivity of my results based on the work of previous researchers (Chang et al., 2006; Llanwarne et al., 2013; Ricci-Cabello et al., 2018). In one specification (Appendix Table 7 Panel 1) I add a dummy variable that equals 1 if the doctor is affiliated with more than one hospital, to account for a variation in quality data reporting due to administrative oversight. The idea is that the data reporting is out of the direct control of the doctor in a hospital, and the more hospitals to which a doctor belongs, the more likely administrative staff are in charge of quality reporting and Medicare paperwork, which may falsely elevate the scores because the staff is specifically hired for their reporting expertise. Alternatively, a doctor of high quality may be sought out by local hospitals and therefore affiliated with more than one hospital.

In addition, I estimate a second regression (Appendix Table 7 Panel 2) that replaces the doctor gender dummy fixed effect with a gender dummy variable that equals one if the patient gender majority and doctor gender are both male. I interacted the clinician gender with average patient gender majority to determine if a score pattern existed between patients and clinicians of the same sex. This variable captures any residual bias in communication between patient and doctor of the same sex.

A third specification (Appendix Table 7 Panel 3) includes the average patient age per provider in the sample to account for age-based stereotypes and generational differences that were not picked up in provider experience and average patient risk score. Ultimately, I chose to exclude these variables from the final reported regression to avoid the risk of an overparameterized model. Additionally, such variables that were on the group level rather than the patient level did not make sense to bring into this model and introduce more uncertainty. Ideally, micro-data on individual clinicians or patients would reduce the amount of uncertainty in the results.

Moreover, I run regressions using the quality category MIPS score instead of the final MIPS score to distill a more accurate reading of the clinician’s technical quality, avoiding the data processing steps that went into calculating the final MIPS score (see Appendix 2 for a description of steps). I also run alternate regressions using either the overall CAHPS score or the mean CAHPS score of selected physician-specific measures in place of the singular measure of the "Patients' Overall Rating of a Provider" (see Table 8 for included measures). The varied levels of specificity of the CAHPS score in my alternate analyses allows me to evaluate how the patient-perceived rating of a provider varies depending on the included survey measures. The results for the alternate regressions are shown in the Appendix in Table 4, Table 5, and Table 6.

6. Results

6.1 Sample Characteristics

The summary statistics are shown in Table 1. The sample distributions are shown as figures in the appendix. A total of 26,841 physicians had MIPS reports and CAHPS scores available for the group(s) to which they belonged. The sample size varied due to missing variable data points for some clinicians. There were no statistically significant differences in mean scores for final MIPS scores between the sample and the CMS dataset at large, and demographic and clinical characteristics in the sample were similar.

The experience of clinicians, as measured by the number of years since their medical school graduation is almost normally distributed around the mean of 20.8 years (see also Figure 6). About 66.7% of clinicians were male (see also Figure 9), and about half of the clinicians were affiliated with more than one hospital. In the sample, 1.1% of clinicians were affiliated with more than one organization (see Section 5 for details about how this was resolved). The mean age of the patients in the sample was 70.7 years; 43% were male and the average risk score was 1.74. The average patient risk was positively skewed and leptokurtic (Figure 7) with a range from 0.49 to 9.76. Risk scores generally range from 0.90 to 1.70, with a score below 1 being considered “relatively healthy” (Pope et al., 2011; Yeatts & Sangvai, 2016). The average age of the beneficiaries per unique clinician was normally distributed within the expected age range of Medicare patients, between eligibility at 65 and average mortality at 80 (Figure 8).

The final MIPS score distribution was clustered at the high end of the range, between scores of 80 and 100 (Figure 4). The mean of the final MIPS score was 91.9. The quality category MIPS score had fewer observations, so the density of scores at the lower end of the range gained more weight in the overall distribution. However, the quality category scores are still clustered near 100 (Figure 5). The high MIPS scores align with the idea that doctors and other clinicians are allowed to choose their MIPS measures. As a result, the MIPS score shows the quality of a doctor in only the areas of practice that the doctor wishes to be visible.

The mean quality category MIPS score was 92.2. The patient-perceived rating measure of the CAHPS survey was available for 2,583 unique clinicians, who constituted the main analytic sample. The patient-perceived rating measure from CAHPS ranged from a low score of 88 to a high of 95 (Figure 3). The mean CAHPS score of patient-perceived rating measure was 92.6. The narrow score range implies that patients were likely to give high scores for a broad question, such as “overall rating of the clinician.” Additionally, clinicians who were more likely to receive a low score may have opted to not report CAHPS survey results, thus eliminating the observations from the sample. The spread of states in the data can be seen in Figure 11.

6.2 Regression Results

I find that, on average, a 1-point increase in the final MIPS score is associated with a 0.003-point increase in the patients' rating of the provider, holding all other variables constant (Table 2). The coefficient, which gets bigger with controls, is positive and statistically significant ($p < 0.001$). However, when I divide the coefficient by the mean CAHPS score, I find that the association between the two scores is negligible. A change of 0.003 of a point is imperceptible relative to the mean CAHPS score of 92.4. The average patient risk score is statistically significant and positive in the regression without fixed effects (Table 2, Column 2), but the variable loses significance when all fixed effects are added to the equation (Table 2, Column 4). This result implies that the fixed effects, likely the state fixed effects, account for the demographic changes in the data set that are determinants of health and patient satisfaction ratings. This result makes sense because state fixed effects generally control for socioeconomic status of the geographic location.

I then found the association between the mean CAHPS score of patient-perceived rating and the final MIPS score separately for each broad specialty type, holding other variables constant and including state and gender fixed effects (Table 3). The two quality score variables were statistically correlated when examining all specialties combined and for internal medicine. The mean of the CAHPS score for all specialties combined was 92.4, with the individual specialty scores each falling between a narrow range of 92.0 and 93.0. The results for all specialties combined, seen in the first column of Table 3, was statistically significant with an estimated coefficient of 0.003 ($p < 0.001$). The correlation between the two scores for clinicians who are in internal medicine was statistically significant ($p < 0.01$), however, the estimated coefficient of 0.005 represents less than 0.01% of the overall magnitude of the CAHPS score for this sample, 92.53. These results further suggest that the final MIPS score has only a weak association to the CAHPS score and that one specialty, namely, internal medicine, is driving the correlation in the results of all specialties combined (Table 3, Column 1).

6.3 Robustness Checks

Below I outline the characteristics of the alternate MIPS and CAHPS variables used in robustness checks along with a brief description of the results. The overall scores for the CAHPS survey were available for 26,123 clinicians who had final MIPS scores (Table 6). The overall CAHPS score ranged from 20 to 99, with greater density in scores ranging from 85 to 99 (Figure 1). The scores for the selected clinician-quality measures of the CAHPS survey were available for 14,609 unique clinicians with a final MIPS score and 14,165 clinicians with a Quality Category Score (Table 5). The mean selected CAHPS score had the same range as the overall CAHPS score but a different distribution; the selected CAHPS scores coalesced around a low score of 20, a score of 60, and a high score of 90 (Figure 2). In addition, the majority of the clinician sample had a Medicare patient base that was a majority a different sex from the clinician (Figure 10).

As the levels for the dependent variable of the CAHPS score became more specific, in other words, included fewer measures, observations were excluded, and the distribution of scores naturally narrowed. The stark narrowing of score ranges implies that patients were more likely to give low scores for more specific aspects of a physician's performance, such as "How Well Providers Communicate" than for an overall perception of the physician. Additionally, physicians who were more likely to receive a low score may have opted to not report CAHPS survey results, thus eliminating the observations from the sample.

Table 1 - Summary Statistics of Overall Sample

	Mean	Standard Deviation	Sample Count
Final MIPS Score	91.914	19.60	26,123
Quality Category MIPS Score	92.158	17.20	25,434
CAHPS Score	92.580	1.83	2,583
Average Patient Age	70.732	4.53	26,204
Average Share of Male Patients	0.428	0.13	24,121
Average Patient Risk Score	1.737	0.78	26,204
Average MD Experience in Years	20.811	10.84	26,772
Male MD (Proportion)	0.667	0.47	26,841
Physician Affiliated with more than one hospital (Proportion)	0.504	0.50	26,841
Physician Affiliated with more than one group (Proportion)	0.011	0.11	26,841

Table 2: Association between the CAHPS Score and the Final MIPS Score

	Base	Only Controls	Only FE	Controls & FE
Final MIPS Score	0.002 (0.002)	0.002 (0.002)	0.003*** (0.001)	0.003*** (0.001)
Provider experience in years		-0.002 (0.003)		0.002 (0.002)
Average patient risk score		0.413*** (0.044)		0.014 (0.021)
State Fixed Effects	No	No	Yes	Yes
Specialty Fixed Effects	No	No	Yes	Yes
Gender Dummy	No	No	Yes	Yes
Mean of Dep. Variable	92.430	92.421	92.410	92.402
Adj. R-Squared	0.000	0.037	0.836	0.837
Observations	2,378	2,336	2,297	2,257

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$
Note: The dependent variable is the singular CAHPS measure titled "Overall rating of provider." This measure is intended to isolate and capture a patient's perception of the quality of his or her doctor. The main independent variable of interest is the final MIPS score, which directly determines reimbursement payments to physicians.

Table 3 - Association between the CAHPS score and the Final MIPS Score by Broad Specialty

	All	Internal Medicine	Cardiology	Primary Care	Specialty Medicine	Surgery	Nursing
Final MIPS Score	0.00341***	0.00504**	-0.00197	0.00334	0.00107	0.00164	0.00142
	(0.001)	(0.002)	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)
Adjusted R^2	0.8338	0.7918	0.7991	0.8724	0.9206	0.9183	0.8652
Mean Dep. Var.	92.41	92.53	92.62	92.08	93.06	92.19	92.01
SD of Dep. Var.	1.780	1.700	1.927	1.785	1.777	1.703	1.671
Observations	2297	707	257	327	155	317	242

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Note: Mean dependent variable and standard deviation of dependent variable shown in lower panel. This model includes state and provider gender fixed effects. "All" includes 11 broad specialties: Internal medicine, Cardiology, Primary care, Specialty medicine, Oncology, Surgery, Radiology, Psychiatry, Nursing, Intensive medicine, and Other allied health.

7. Discussion

7.1 Interpretation of Results

The relationship between the patients' rating of the provider (CAHPS score) and the provider's MIPS score is statistically significant but is not economically significant. As a result, patient-perceived quality ratings would be a poor substitute for ratings of clinical quality. This finding is in-line with existing literature, which suggests that both scores are important to capture a more complete view of a physician's "true" quality (Anhang Price et al., 2014). However, previous studies use datasets from survey systems other than AHRQ CAHPS (Bowling et al., 2012; Rao M. et al., 2006).

There are many limitations in the current dataset. The national benchmarking process for determining final MIPS score deciles is based on annual averages from that year's pool of Medicare enrolled physicians. As a result, researchers may find it troublesome to measure year-to-year comparisons of a clinician's performance. Bonus points added on top of the raw performance score also obscure the physician's clinical quality performance and further hinder time-series comparison capabilities. This study is limited to a cross-sectional analysis on data from 2017. Moreover, researchers may be unable to capture certain underlying characteristics of a physician if the physician chooses to only report measures that he/she already excels in or that are inexpensive to improve upon.

A similar issue arises in the CAHPS data, as researchers are unable to observe micro-data on unique patient rating scores for each physician within a group practice. Subsequently, in order to connect patient experience to clinical quality, I necessarily assume that all providers within a group have the same patient ratings. This assumption is logical in that providers likely know each other and will choose to affiliate with other providers of like quality. In addition, it could be difficult to capture what exactly could improve patients' experiences of care because each patient has his or her own rating scale. If these scores are not normalized, mis-anchoring of scores could occur from patient to patient for one physician. The utility of the CAHPS scores for research is also limited because of the manner in which the scores are reported. CAHPS scores are available to the public as top-box scores, making the marginal analysis of quality between numeric scores impossible.

7.2 General Policy Implications

The limitations of this study bring up three policy implications regarding MIPS and CAHPS, which include public reporting of quality scores, physician reimbursement, and patient welfare. For one, policymakers must decide if both or either of the individual physician quality scores should be publicly reported and thus visible to all patients. Currently, only group-level physician quality data is available. If both scores become publicly available, more information about the policy and the score calculations should also become more accessible. It is important that patients understand what comprises the different scores and how to interpret the final composite number. The information should clarify that neither technical quality nor patient-perceived quality rating is inherently more important than the other. Second, because the MIPS program is budget neutral, there is a likelihood that the physician reimbursement model will have to change as more physicians enroll. Some research has shown that MIPS may be more of a penalty payment system than it is a system to reward high quality physicians (Young, 2019; Verma, 2018). Lastly, patient review platforms, such as Healthgrades and Yelp, and publicly reported CAHPS and MIPS scores have benefits and drawbacks. Many existing online platforms increase the visibility of individual patient profiles, which could improve accountability to add honest and specific evaluations of the health provider (Bardach et al., 2013; Tran & Lee, 2017). This more personal platform allows clinicians a chance to see more definitive ways to improve patient experiences. On the other hand, CAHPS offers patients anonymity to express negative experiences more openly. CAHPS also is not as strongly subject to response bias and has validated questions (Chaitoff et al., 2017; Fenton et al., 2017; Stern et al., 2012). It would be interesting to see how the two types of patient experience platforms evolve to affect physicians' behavior.

8. Conclusion

This research examines the link between patient-reported quality measures and clinical quality measures in Medicare's recent Quality Payment Program. This study is one of the first to use data from the first year of the policy's implementation, which began in 2017. Additionally, I focus on outpatient settings rather than the more often studied data from inpatient hospitals. I find that the correlation between the patient-perceived rating of a clinician and the technical rating of the clinician is statistically significant but not economically significant. The findings of this study support the results of previous research concerning the correlation between patient quality metrics and clinician technical quality (Girotra et al., 2012; Sequist et al., 2008).

Future research should focus on investigating the reason why clinical scores and patient scores are weakly correlated. Researchers could investigate the mechanism that is driving the size of each quality score or identify features of healthcare that patients value beyond clinical quality. More research should be done to further target what a patient considers in his/her judgement of quality on the CAHPS survey. It would also be valuable to study the characteristics of physicians. For example, does the final MIPS score predict vertical integration or consolidation of certain physician groups? Could the two quality scores reduce adverse selection in the market between hospital "employers" and physicians “employees”? Although determining the drivers behind the weak correlation between MIPS and CAHPS quality scores was beyond the scope of this study, further investigation in this area may better inform patient choice and empower physicians to improve patient experiences through patient-centered and efficient care.

References

Agency for Healthcare Research and Quality (AHRQ). (2018, October). *CAHPS for Merit-Based Incentive Payment System (MIPS) Survey*. http://www.ahrq.gov/cahps/surveys-guidance/cg/cahps-mips.html

American College of Physicians (ACP). *Qpp-macra-explained.pdf*. (n.d.). Retrieved September 7, 2019, from https://www.acponline.org/system/files/documents/practice-resources/business-resources/payment/medicare/macra/qpp-macra-explained.pdf

Ancker, J. S., Singh, M. P., Thomas, R., Edwards, A., Snyder, A., Kashyap, A., & Kaushal, R. (2013). Predictors of success for electronic health record implementation in small physician practices. *Applied Clinical Informatics*, *4*(1), 12–24. https://doi.org/10.4338/ACI-2012-09-RA-0033

Anhang Price, R., Elliott, M. N., Zaslavsky, A. M., Hays, R. D., Lehrman, W. G., Rybowski, L., Edgman-Levitan, S., & Cleary, P. D. (2014). Examining the role of patient experience surveys in measuring health care quality. *Medical Care Research and Review: MCRR*, *71*(5), 522–554. https://doi.org/10.1177/1077558714541480

Baim-Lance, A., Tietz, D., Schlefer, M., & Agins, B. (2016). Health Care User Perspectives on Constructing, Contextualizing, and Co-Producing "Quality of Care." *Qualitative Health Research*, *26*(2), 252–263. https://doi.org/10.1177/1049732315569736

Bardach, N. S., Asteria-Peñaloza, R., Boscardin, W. J., & Dudley, R. A. (2013). The relationship between commercial website ratings and traditional hospital performance measures in the USA. *BMJ Quality & Safety*, *22*(3), 194–202. https://doi.org/10.1136/bmjqs-2012-001360

Becker, G. (2007). Health as Human Capital: Synthesis and Extensions. *Oxford Economic Papers, 59*(3), new series, 379-410. Retrieved April 27, 2020, from www.jstor.org/stable/4500116

Boccuti, C., Fields, C., Casillas, G., & Hamel, L. (2015, October 30). Primary Care Physicians Accepting Medicare: A Snapshot. *The Henry J. Kaiser Family Foundation*. https://www.kff.org/medicare/issue-brief/primary-care-physicians-accepting-medicare-a-snapshot/

Bowling, A., Rowe, G., Lambert, N., Waddington, M., Mahtani, K. R., Kenten, C., Howe, A., & Francis, S. A. (2012). The measurement of patients' expectations for health care: A review and psychometric testing of a measure of patients' expectations. *Health Technology Assessment (Winchester, England)*, *16*(30), i–xii, 1–509. https://doi.org/10.3310/hta16300

Braveman, P., & Gottlieb, L. (2014). The Social Determinants of Health: It's Time to Consider the Causes of the Causes. *Public Health Reports, 129*(Suppl 2), 19–31.

Bresnahan, T., & Levin, J. (2012). *Vertical Integration and Market Structure* (Discussion Paper No. 11–010). Stanford Institute for Economic Policy Research. https://econpapers.repec.org/paper/sipdpaper/11-010.htm

Burwell SM. Setting Value-Based Payment Goals — HHS Efforts to Improve U.S. Health Care. N Engl J Med. 2015;372:897–899. doi: 10.1056/NEJMp1500445

Centers for Medicare & Medicaid Services. (n.d.). *QPP Overview—QPP*. Quality Payment Program Overview. Retrieved September 7, 2019, from https://qpp.cms.gov/about/qpp-overview

Chaitoff, A., Sun, B., Windover, A., Bokar, D., Featherall, J., Rothberg, M. B., & Misra-Hebert, A. D. (2017). Associations Between Physician Empathy, Physician Characteristics, and Standardized Measures of Patient Experience. *Academic Medicine: Journal of the Association of American Medical Colleges*, *92*(10), 1464–1471. https://doi.org/10.1097/ACM.0000000000001671

Chee, T. T., Ryan, A. M., Wasfy, J. H., & Borden, W. B. (2016). Current State of Value-Based Purchasing Programs. *Circulation*, *133*(22), 2197–2205. https://doi.org/10.1161/CIRCULATIONAHA.115.010268

Chen, J. G., Zou, B., & Shuster, J. (2017). Relationship Between Patient Satisfaction And Physician Characteristics. *Journal of Patient Experience*, *4*(4), 177–184. https://doi.org/10.1177/2374373517714453

Clemens, J., & Gottlieb, J. D. (2017). In the Shadow of a Giant: Medicare's Influence on Private Physician Payments. *The Journal of Political Economy*, *125*(1), 1–39. https://doi.org/10.1086/689772

Clemens, J., Gottlieb, J. D., & Molnár, T. L. (2017). Do health insurers innovate? Evidence from the anatomy of physician payments. *Journal of Health Economics*, *55*, 153–167. https://doi.org/10.1016/j.jhealeco.2017.07.001

Cramer, D., & Howitt, D. L. (2005). *The SAGE dictionary of statistics: A practical resource for students in the social sciences* (3rd ed.). London/Thousand Oaks: Sage, New Delhi.

Curto, V., Einav, L., Finkelstein, A., Levin, J., & Bhattacharya, J. (2019). Health Care Spending and Utilization in Public and Private Medicare. *American Economic Journal. Applied Economics*, *11*(2), 302–332. https://doi.org/10.1257/app.20170295

Drake, K. M., Hargraves, J. L., Lloyd, S., Gallagher, P. M., & Cleary, P. D. (2014). The effect of response scale, administration mode, and format on responses to the CAHPS Clinician and Group survey. *Health Services Research*, *49*(4), 1387–1399. https://doi.org/10.1111/1475-6773.12160

Ehrlich, A., Kostecki, J., & Olkaba, H. (2017). Trends in dermatology practices and the implications for the workforce. *Journal Of The American Academy Of Dermatology*, *77*(4), 746–752. https://doi.org/10.1016/j.jaad.2017.06.030

Espay, A. J., Norris, M. M., Eliassen, J. C., Dwivedi, A., Smith, M. S., Banks, C., Allendorfer, J. B., Lang, A. E., Fleck, D. E., Linke, M. J., & Szaflarski, J. P. (2015). Placebo effect of medication cost in Parkinson disease. *Neurology*, *84*(8), 794–802. https://doi.org/10.1212/WNL.0000000000001282

Everitt, B. S. (2002). *The Cambridge dictionary of statistics* (2nd ed.). Cambridge, UK: Cambridge University Press.

Fenton, J. J., Jerant, A., Kravitz, R. L., Bertakis, K. D., Tancredi, D. J., Magnan, E. M., & Franks, P. (2017). Reliability of Physician-Level Measures of Patient Experience in Primary Care. *Journal of General Internal Medicine*, *32*(12), 1323–1329. https://doi.org/10.1007/s11606-017-4175-y

Fowler, F. J. (1997). *Choosing questions to measure the quality of experience with medical care providers and health care plans.* Proceedings of the Section on Survey Research Methods. 1997, 51-54.

Girotra, S., Cram, P., & Popescu, I. (2012). Patient satisfaction at America's lowest performing hospitals. *Circulation. Cardiovascular Quality and Outcomes*, *5*(3), 365–372. https://doi.org/10.1161/CIRCOUTCOMES.111.964361

Hanefeld, J., Powell-Jackson, T., & Balabanova, D. (2017). Understanding and measuring quality of care: Dealing with complexity. *Bulletin of the World Health Organization*, *95*(5), 368–374. https://doi.org/10.2471/BLT.16.179309

Hargraves, J. L., Cosenza, C., Elliott, M. N., & Cleary, P. D. (2019). The effect of different sampling and recall periods in the CAHPS Clinician & Group (CG-CAHPS) survey. *Health Services Research*, *54*(5), 1036–1044. https://doi.org/10.1111/1475-6773.13173

Harris-Kojetin LD, Fowler FJ Jr, Brown JA, Schnaier JA, Sweeny SF. *The use of cognitive testing to develop and evaluate CAHPS 1.0 core survey items.* Med Care. 1999 Mar;37(3 Suppl):MS10-21.

Hays, R. D., Mallett, J. S., Haas, A., Kahn, K. L., Martino, S. C., Gaillot, S., & Elliott, M. N. (2018). Associations of CAHPS Composites With Global Ratings of the Doctor Vary by Medicare Beneficiaries' Health Status. *Medical Care, 56(8), 736–739.* Retrieved January 28, 2020, from http://web.b.ebscohost.com/ehost/detail/detail?vid=0&sid=decb7466-b76c-4565-8723-4aba36fd1b35%40sessionmgr102&bdata=JkF1dGhUeXBlPWlwLHNzbyZzaXRlPWVob3N0LWxpdmUmc2NvcGU9c2l0ZQ%3d%3d#AN=29939911&db=cmedm

Health Care Payment Learning & Action Network. (2018, October 22). *Measuring Progress: Adoption of Alternative Payment Models in Commercial, Medicaid, Medicare Advantage, and Medicare Fee-for-Service Programs.* 2018 APM Measurement Effort. https://hcp-lan.org/2018-apm-measurement/

Iezzoni, L. I. (1994). Using risk-adjusted outcomes to assess clinical practice: An overview of issues pertaining to risk adjustment. *The Annals of Thoracic Surgery*, *58*(6), 1822–1826. https://doi.org/10.1016/0003-4975(94)91721-3

James, J. (2012, October 11). *Health Policy Brief: Pay-for-Performance*. Health Affairs. https://www.healthaffairs.org/do/10.1377/hpb20121011.90233/full/

Kash, B., & Tan, D. (2016). Physician Group Practice Trends: A Comprehensive Review. *Journal of Hospital & Medical Management*, *2*(1). https://doi.org/10.4172/2471-9781.100008

Klein, D. J., Elliott, M. N., Haviland, A. M., Saliba, D., Burkhart, Q., Edwards, C., & Zaslavsky, A. M. (2011). Understanding Nonresponse to the 2007 Medicare CAHPS Survey. *Gerontologist*, *51*(6), 843–855.

Koch, T. G., Wendling, B. W., & Wilson, N. E. (2017). How vertical integration affects the quantity and cost of care for Medicare beneficiaries. *Journal of Health Economics*, *52*, 19–32. https://doi.org/10.1016/j.jhealeco.2016.12.007

Levitt, T. (1981, May 1). Marketing Intangible Products and Product Intangibles. *Harvard Business Review*, *May 1981*. https://hbr.org/1981/05/marketing-intangible-products-and-product-intangibles

Llanwarne, N. R., G. A. Abel, M. N. Elliott, C. A. Paddison, G. Lyratzopoulos, J. L. Campbell, and M. Roland. 2013. “Relationship between Clinical Quality and Patient Experience: Analysis of Data from the English Quality and Outcomes framework and the National GP Patient Survey.” *Annals of Family Medicine* 11 (5): 467–72.

Manning, W. G., Newhouse, J. P., Duan, N., Keeler, E. B., Leibowitz, A., & Marquis, M. S. (1987). Health insurance and the demand for medical care: Evidence from a randomized experiment. *The American Economic Review*, *77*(3), 251–277.

Martiniano R., & Moore, J. (February 2018). *Health Care Employment Projections, 2016-2026: An Analysis of Bureau of Labor Statistics Projections by Setting and by*

Occupation. Rensselaer, NY: Center for Health Workforce Studies, School of Public Health, SUNY Albany;

Martino, S. C., Shaller, D., Schlesinger, M., Parker, A. M., Rybowski, L., Grob, R., Cerully, J. L., & Finucane, M. L. (2017). CAHPS and Comments: How Closed-Ended Survey Questions and Narrative Accounts Interact in the Assessment of Patient Experience. *Journal of Patient Experience*, *4*(1), 37–45. https://doi.org/10.1177/2374373516685940

Mehta, S. J. (2015). Patient Satisfaction Reporting and Its Implications for Patient Care. *AMA Journal of Ethics*, *17*(7), 616–621. https://doi.org/10.1001/journalofethics.2015.17.7.ecas3-1507.

Oster, E. (2012). HIV and sexual behavior change: Why not Africa? *Journal of Health Economics*, *31*(1), 35–49. https://doi.org/10.1016/j.jhealeco.2011.12.006

Owaid, O. (2017). The Death of Private Practice: How the Rising Cost of Healthcare is Destroying Physician Autonomy. *Brooklyn Journal of Corporate, Financial & Commercial Law*, *11*(2), 521.

Peck, B. M. (2011). Age-Related Differences in Doctor-Patient Interaction and Patient Satisfaction. *Current Gerontology and Geriatrics Research*, *2011*. https://doi.org/10.1155/2011/137492

Poot, A. J., Wopereis, D. M., den Elzen, W. P. J., Gussekloo, J., & Blom, J. W. (2019). Changes in patient satisfaction related to their perceived health state during implementation of improved integrated care for older persons. *PloS One*, *14*(5), e0216028. https://doi.org/10.1371/journal.pone.0216028

Pope, G. C., Kautter J., Ingber, M. J., Freeman S., Sekar R., & Newhart, C. (2011). *Evaluation of the CMS-HCC Risk Adjustment Model, Final Report*. 127.

Quigley, B. M., Sova, C. C., Brenner, D. M., Keefer, L. A., Sitrin, M. D., Radziwon, C. D., Krasner, S. S., & Lackner, J. M. (2018). (Can't Get No) Patient Satisfaction: The Predictive Power of Demographic, GI, and Psychological Factors in IBS Patients. *Journal of Clinical Gastroenterology*, *52*(7), 614–621. https://doi.org/10.1097/MCG.0000000000000906

Rao, M., Clarke, A., Sanderson, C., & Hammersley, R. (2006). Patients' own assessments of quality of primary care compared with objective records based measures of technical quality of care: Cross sectional study. *BMJ (Clinical Research Ed.)*, *333*(7557), 19. https://doi.org/10.1136/bmj.38874.499167.7C

Rao, S. K., Kimball, A. B., Lehrhoff, S. R., Hidrue, M. K., Colton, D. G., Ferris, T. G., & Torchiana, D. F. (2017). The Impact of Administrative Burden on Academic Physicians: Results of a Hospital-Wide Physician Survey. *Academic Medicine: Journal of the Association of American Medical Colleges*, *92*(2), 237–243. https://doi.org/10.1097/ACM.0000000000001461

Rao, S. R., Desroches, C. M., Donelan, K., Campbell, E. G., Miralles, P. D., & Jha, A. K. (2011). Electronic health records in small physician practices: Availability, use, and perceived benefits. *Journal Of The American Medical Informatics Association: JAMIA*, *18*(3), 271–275. https://doi.org/10.1136/amiajnl-2010-000010

Ricci-Cabello, I., Stevens, S., Dalton, A. R. H., Griffiths, R. I., Campbell, J. L., & Valderas, J. M. (2018). Identifying Primary Care Pathways from Quality of Care to Outcomes and Satisfaction Using Structural Equation Modeling. *Health Services Research*, *53*(1), 430–449. https://doi.org/10.1111/1475-6773.12666

Ross, HJ. (2014). *Everyday Bias: Identifying and Navigating Unconscious Judgments in Our Daily Lives*. Lanham, MD: Rowman & Littlefield Publishers.

Scholle, S. H., Vuong, O., Ding, L., Fry, S., Gallagher, P., Brown, J. A., Hays, R. D., & Cleary, P. D. (2012). Development of and field test results for the CAHPS PCMH Survey. *Medical Care*, *50 Suppl*, S2-10. https://doi.org/10.1097/MLR.0b013e3182610aba

Sequist, T.D., Schneider, E.C., Anastario, M., Odigie, E.G., Marshall, R., Rogers, W.H., et al. (2008). Quality monitoring of physicians: linking patients' experiences of care to clinical quality and outcomes. *Journal of General Internal Medicine*, 23(11):1784–1790.

Sharp, B., Johnson, J., Hamedani, A. G., Hakes, E. B., & Patterson, B. W. (2019). What Are We Measuring? Evaluating Physician-Specific Satisfaction Scores Between Emergency Departments. *The Western Journal of Emergency Medicine*, *20*(3), 454–459. https://doi.org/10.5811/westjem.2019.4.41040

Stern, R. J., Fernandez, A., Jacobs, E. A., Neilands, T. B., Weech-Maldonado, R., Quan, J., Carle, A., & Seligman, H. K. (2012). Advances in measuring culturally competent care: A confirmatory factor analysis of CAHPS-CC in a safety-net population. *Medical Care*, *50*(9 Suppl 2), S49-55. https://doi.org/10.1097/MLR.0b013e31826410fb

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*, 34–42.

The Henry J. Kaiser Family Foundation. (2020, February 21). KFF Health Tracking Poll – February 2020: Health Care in the 2020 Election - Methodology. *The Henry J. Kaiser Family Foundation*. https://www.kff.org/report-section/kff-health-tracking-poll-february-2020-health-care-in-the-2020-election-methodology/

Tran, N. N., & Lee, J. (2017). Online Reviews as Health Data: Examining the Association Between Availability of Health Care Services and Patient Star Ratings Exemplified by the Yelp Academic Dataset. *JMIR Public Health and Surveillance*, *3*(3), e43. https://doi.org/10.2196/publichealth.7001

Verma, S. (2018, November 8). *Quality Payment Program (QPP) Year 1 Performance Results | CMS*. https://www.cms.gov/blog/quality-payment-program-qpp-year-1-performance-results

Woolhandler, S., & Himmelstein, D. U. (2014). Administrative work consumes one-sixth of U.S. physicians' working hours and lowers their career satisfaction. *International Journal of Health Services: Planning, Administration, Evaluation*, *44*(4), 635–642. https://doi.org/10.2190/HS.44.4.a

Writers, S. (2019, November 6). HCPCS Codes in Billing and Coding. Retrieved November 11, 2019, from MedicalBillingandCoding.org website: https://www.medicalbillingandcoding.org/hcpcs-codes/

Yeatts, J. P., & Sangvai, D. (2016). HCC Coding, Risk Adjustment, and Physician Income: What You Need to Know. *Family Practice Management*, *23*(5), 24–27.

Young, K. D. (2019, December 12). *MedPAC Chief Expects Waning Support for Medicare's MIPS*. Medscape. http://www.medscape.com/viewarticle/922509

Zgierska, A., Miller, M., & Rabago, D. (2012). Patient satisfaction, prescription drug abuse, and potential unintended consequences. *JAMA*, *307*(13), 1377–1378. https://doi.org/10.1001/jama.2012.419

Zgierska, A., Rabago, D., & Miller, M. M. (2014). Impact of patient satisfaction ratings on physicians and clinical care. *Patient Preference and Adherence*, *8*, 437–446. https://doi.org/10.2147/PPA.S59077

Zhang, X., Lin, D., Pforsich, H., & Lin, V. W. (2020). Physician workforce in the United States of America: Forecasting nationwide shortages. *Human Resources for Health*, *18*(1), 8. https://doi.org/10.1186/s12960-020-0448-3

Appendix

Figure 1

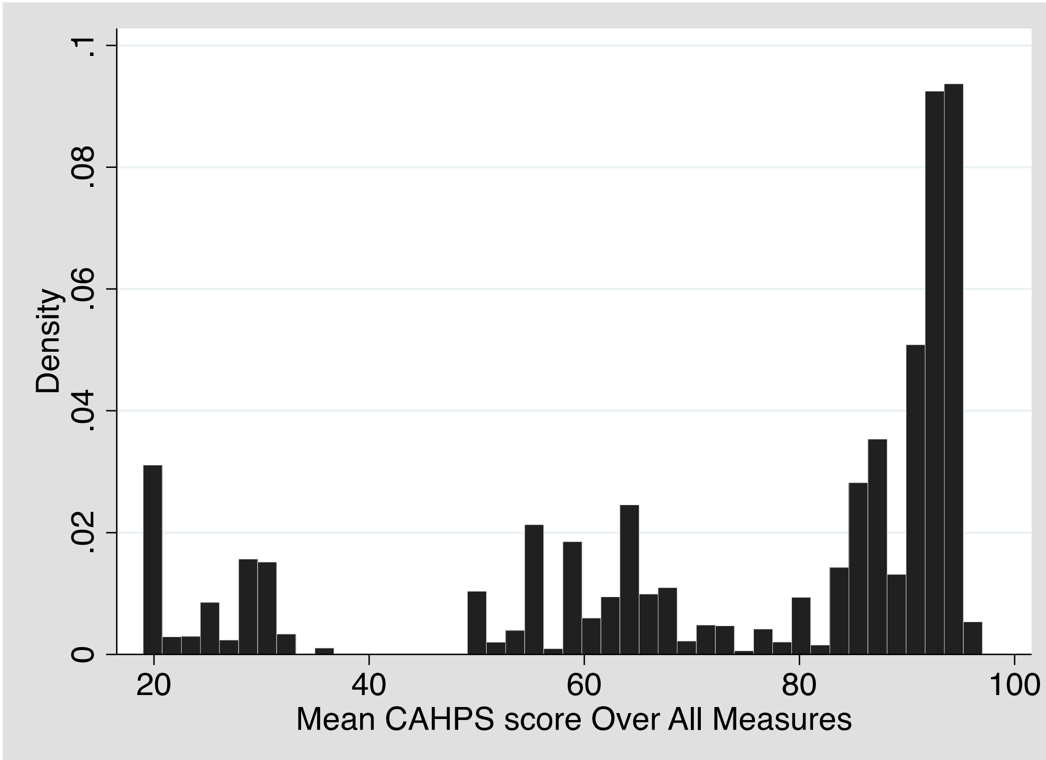


Figure 2

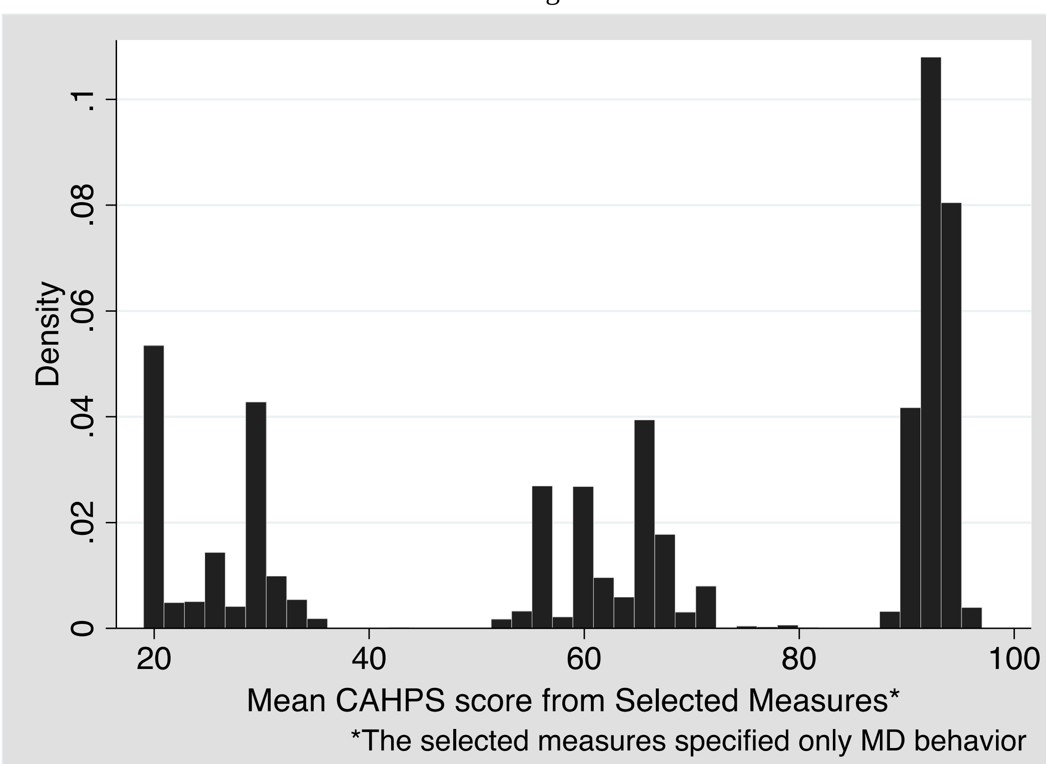


Figure 3

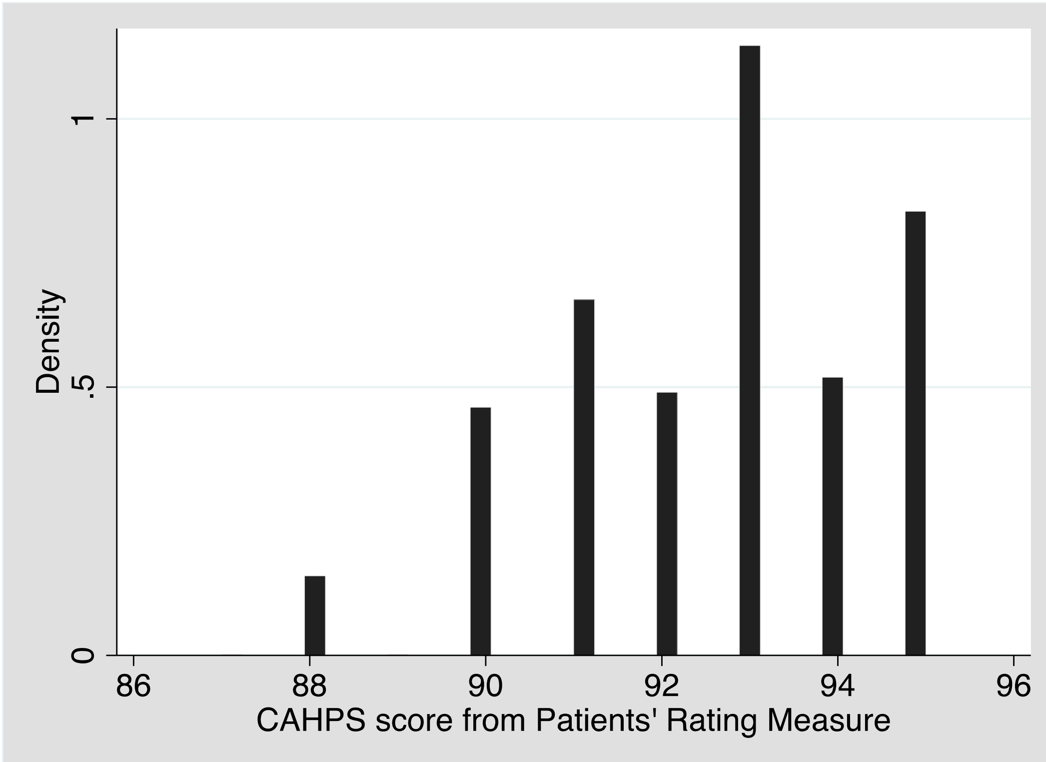


Figure 4

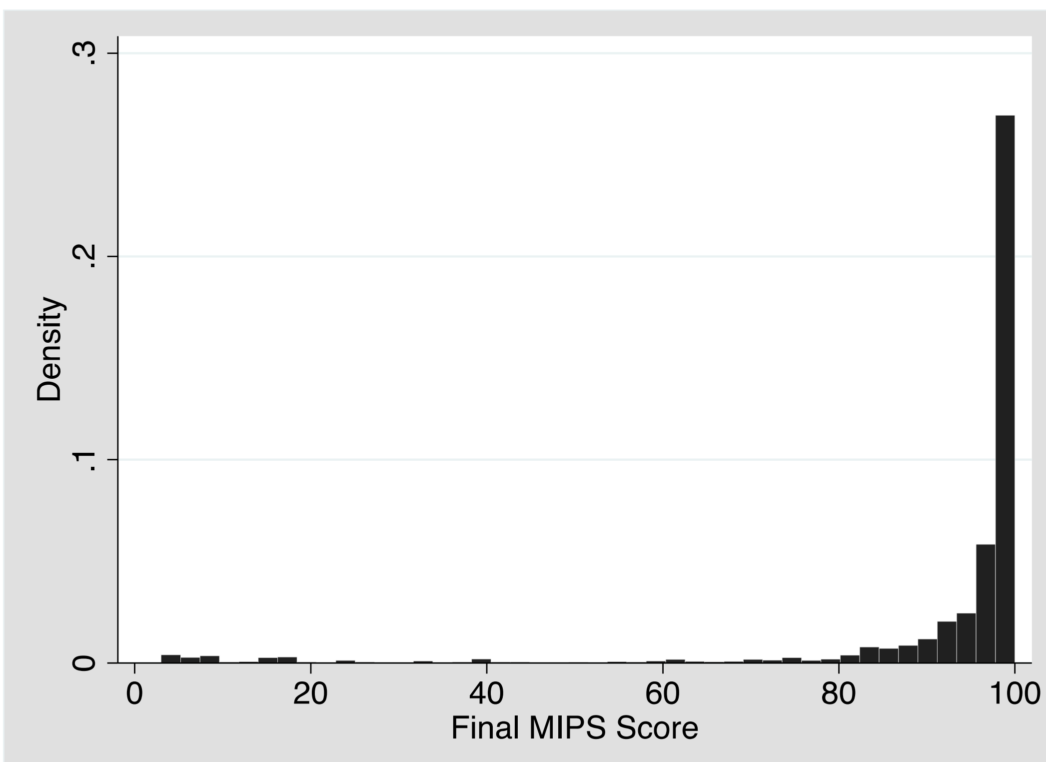


Figure *5*

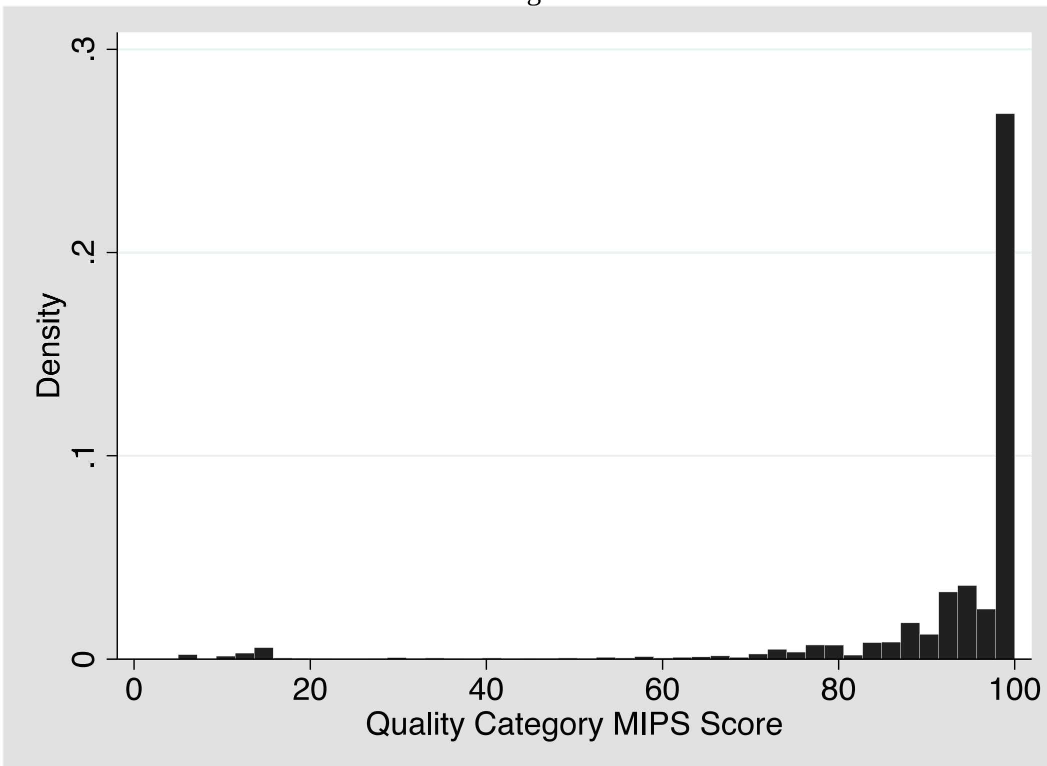


Figure 6

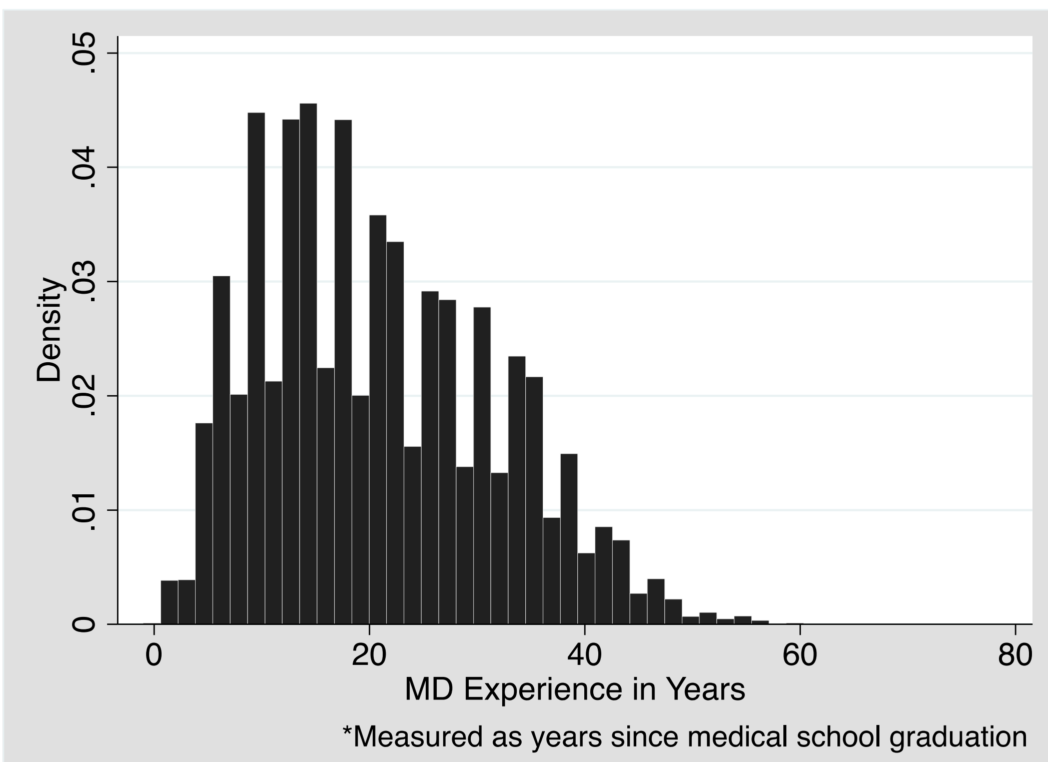


Figure 7

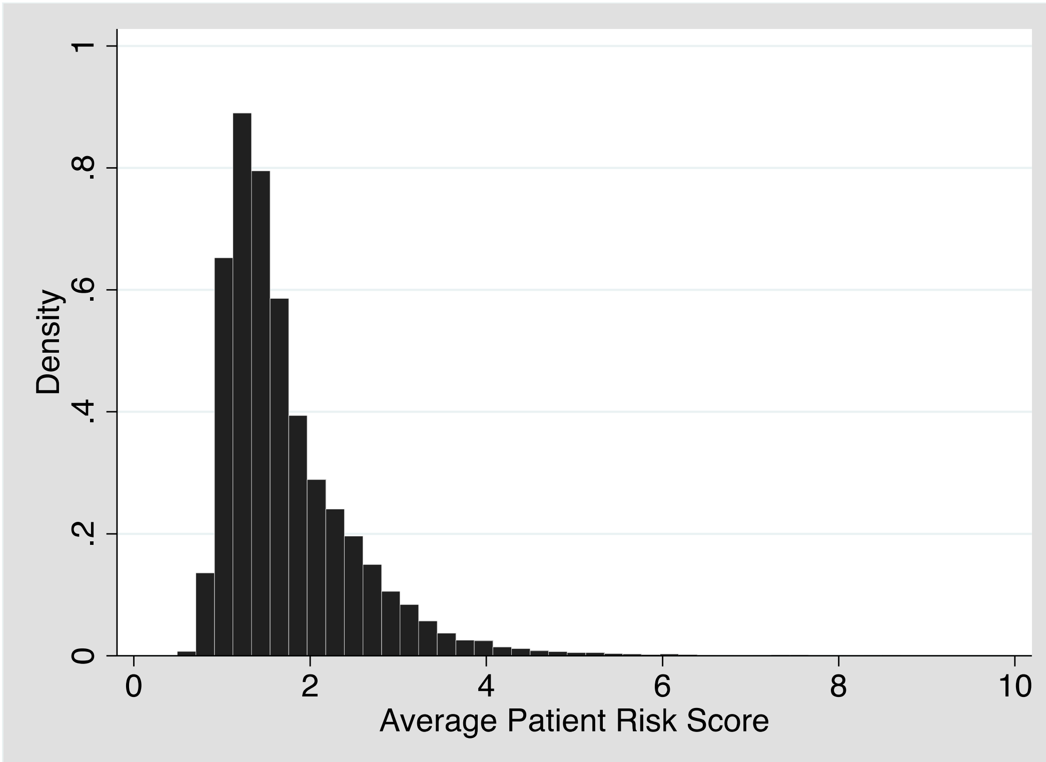


Figure 8

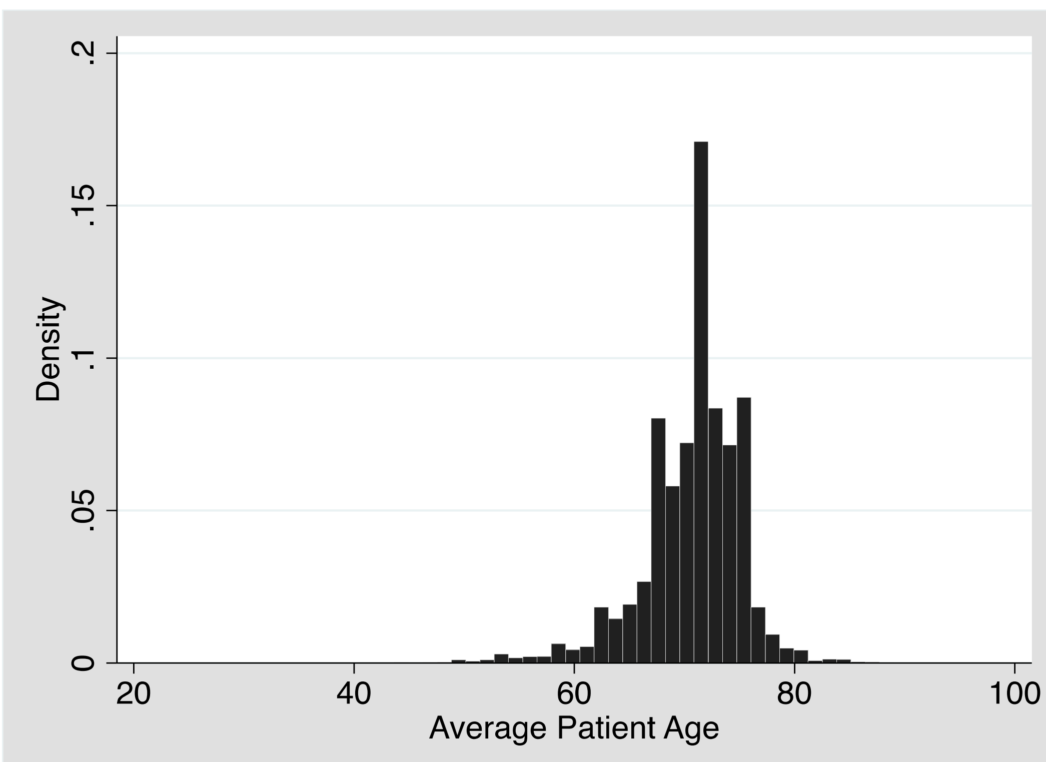


Figure 9

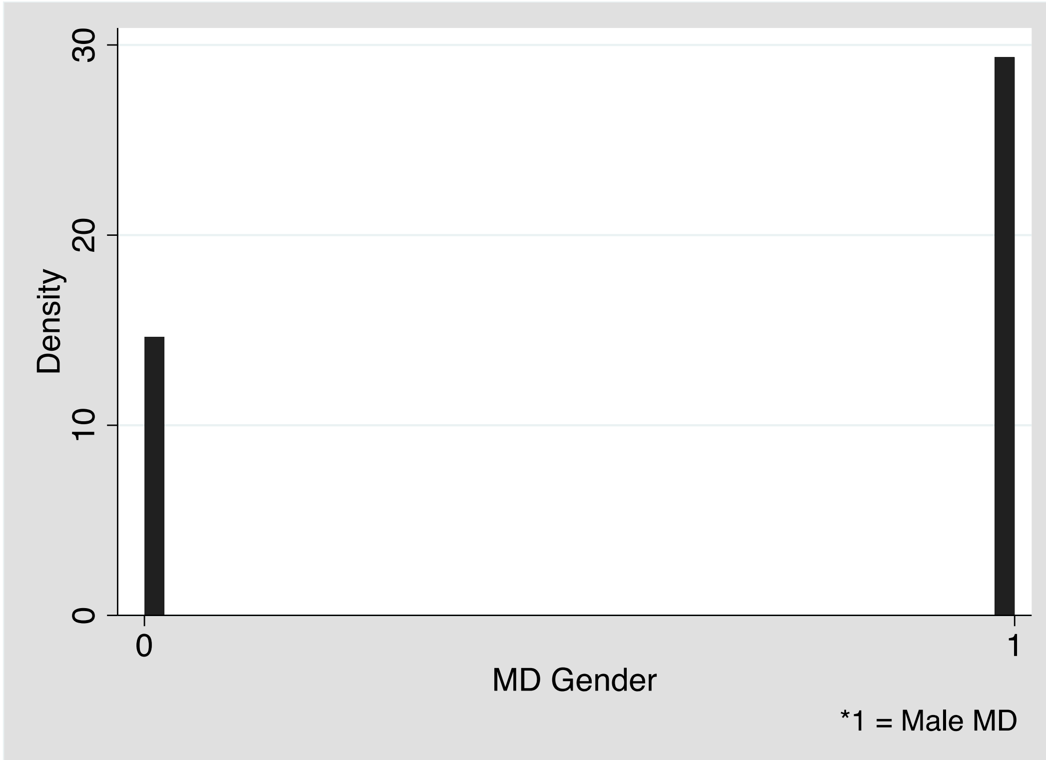


Figure 10

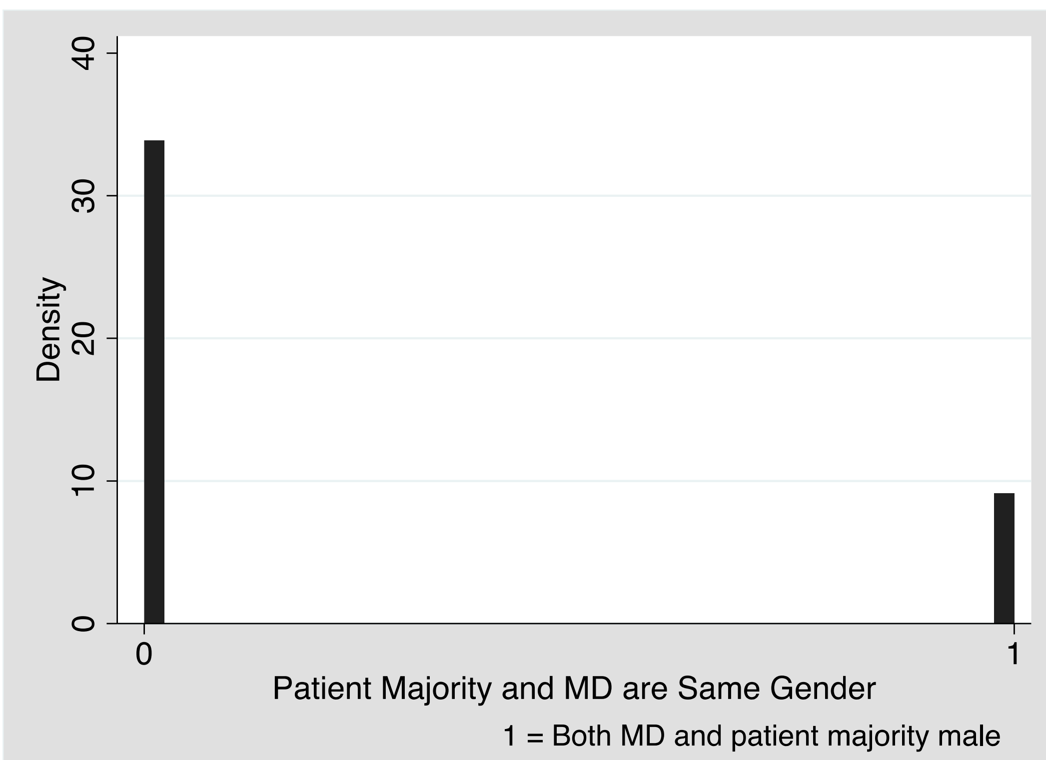
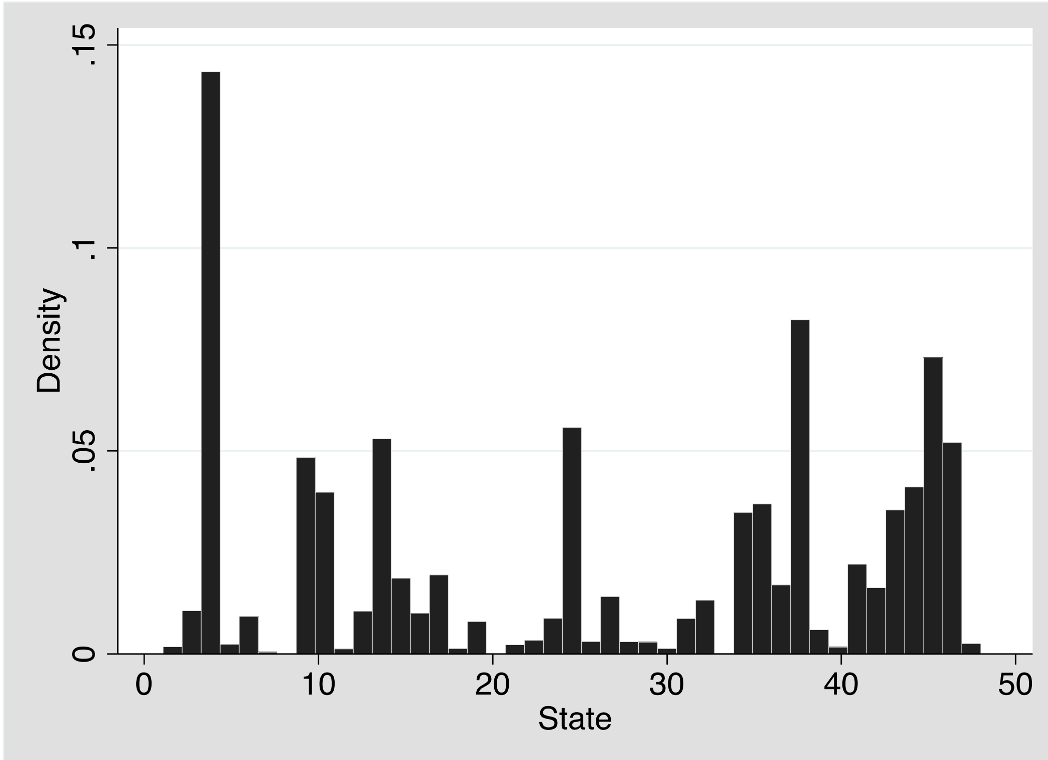


Figure 11



Below are the regression result tables from alternate model specifications, with Table 2 repeated for ease of comparison. The regressions adjust three levels of specificity for the dependent variable and two levels of specificity for the main dependent variable.

Table 2: Association between the CAHPS Score and the Final MIPS Score

	Base	Only Controls	Only FE	Controls & FE
Final MIPS Score	0.002 (0.002)	0.002 (0.002)	0.003*** (0.001)	0.003*** (0.001)
Provider experience in years		-0.002 (0.003)		0.002 (0.002)
Average patient risk score		0.413*** (0.044)		0.014 (0.021)
State Fixed Effects	No	No	Yes	Yes
Specialty Fixed Effects	No	No	Yes	Yes
Gender Dummy	No	No	Yes	Yes
Mean of Dep. Variable	92.430	92.421	92.410	92.402
Adj. R-Squared	0.000	0.037	0.836	0.837
Observations	2,378	2,336	2,297	2,257

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$
Note: The dependent variable is the singular CAHPS measure titled “Overall rating of provider.” This measure is intended to isolate and capture a patient’s perception of the quality of his or her doctor. The main independent variable of interest is the final MIPS score, which directly determines reimbursement payments to physicians.

Table 4: Patients' Rating of Provider CAHPS against the Quality Category MIPS Score

	Base	Only FE	Controls & FE
Quality Category Score	-0.002 (0.003)	0.006*** (0.001)	0.006*** (0.001)
Provider experience in years			0.001 (0.002)
Average patient risk score			-0.021 (0.026)
State Fixed Effects	No	Yes	Yes
Specialty Fixed Effects	No	Yes	Yes
Gender Dummy	No	Yes	Yes
Mean of Dep. Variable	92.419	92.419	92.410
Adj. R-Squared	-0.000	0.839	0.839
Observations	2,308	2,308	2,267

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$
Note: This table shows the most specific level of the dependent variable and of the main independent variable of interest. The dependent variable is the singular CAHPS measure titled “Overall rating of provider.” This measure is intended to isolate and capture a patient’s perception of the quality of his or her doctor. The main independent variable of interest is the quality category score, which is one of the four main components of the final MIPS score. The quality category pulls from a smaller pool of performance measures to calculate the score. The measures included in the calculation focus on the actions of a doctor that would signify quality.

Table 5: Mean Selected CAHPS against Quality Category MIPS Score and Final MIPS Score

	Base		Only FE		Controls & FE	
Final MIPS Score	0.099*** (0.011)		0.030** (0.009)		0.032*** (0.009)	
Quality Category Score		0.120*** (0.013)		0.026* (0.011)		0.031** (0.011)
Provider experience in years					-0.004 (0.019)	-0.004 (0.019)
Average patient risk score					2.331*** (0.308)	2.450*** (0.316)
State Fixed Effects	No	No	Yes	Yes	Yes	Yes
Specialty Fixed Effects	No	No	Yes	Yes	Yes	Yes
Gender Dummy	No	No	Yes	Yes	Yes	Yes
Mean of Dep. Variable	65.826	65.615	65.747	65.539	65.826	65.616
Adj. R-Squared	0.005	0.006	0.380	0.378	0.381	0.380
Observations	14,609	14,165	13,916	13,505	14,249	13,819

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$
Note: This table shows the medium specificity level of the dependent variable. The table also shows both levels of specificity for the main independent variable of interest. The dependent variable is the mean score of a group of CAHPS measures that relate to doctor performance rather than performance of the staff or facility managers. This dependent variable is broader than the one measure of the patient's "Overall rating of provider" but more specific than averaging the score of all CAHPS measures.

Table 6: Mean Overall CAHPS against Quality Category MIPS Score & Final MIPS Score

	Base		FE Only		FE & Controls	
Final MIPS Score	0.084*** (0.008)		0.069*** (0.007)		0.068*** (0.007)	
Quality Category Score		0.097*** (0.009)		0.078*** (0.008)		0.077*** (0.008)
Provider experience in years					-0.019 (0.014)	-0.018 (0.014)
Average patient risk score					1.382*** (0.229)	1.486*** (0.233)
State Fixed Effects	No	No	Yes	Yes	Yes	Yes
Specialty Fixed Effects	No	No	Yes	Yes	Yes	Yes
Gender Dummy	No	No	Yes	Yes	Yes	Yes
Mean of Dep. Variable	74.062	73.978	74.062	73.978	74.052	73.972
Adj. R-Squared	0.005	0.005	0.211	0.211	0.215	0.215
Observations	26,123	25,434	26,123	25,434	25,433	24,773

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$
Note: This table shows the broadest specificity level of the dependent variable and show both levels of specificity for the main independent variable of interest. The dependent variable is the mean score of all CAHPS measures in the patient feedback survey, including measures that relate to performance of the staff or facility managers as well as the doctor.

Table 7: Robustness Check
Patients' Rating of Provider CAHPS against the Final MIPS Score

	Panel 1	Panel 2	Panel 3
Final MIPS Score	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)
Provider experience in years	0.002	-0.000	0.003
	(0.002)	(0.002)	(0.002)
Average patient age			-0.021***
			(0.004)
State Fixed Effects	Yes	Yes	Yes
Specialty Fixed Effects	Yes	Yes	Yes
Gender Dummy	Yes	No	Yes
Hospital Proportion	Yes	No	No
Mean of Dep. Variable	92.410	92.364	92.402
Adj. R-Squared	0.837	0.839	0.839
Observations	2,290	2,121	2,257

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$
Note: All regressions have state & specialty FE and provider experience. Panel 1 includes gender FE and a dummy variable that indicates whether the provider affiliated with more than 1 hospital. Panel 2 includes the alternate gender dummy that equals 1 if the majority patients and physician have same gender. Panel 3 includes average patient age as a control variable

Table 8

CAHPS Measure	"Selected Physician-specific Measures"
2	"CAHPS for MIPS SSM: How Well Providers Communicate"
3	"CAHPS for MIPS SSM: Patient's Rating of Provider"
4	"CAHPS for MIPS SSM: Health Promotion and Education"
7	"CAHPS for MIPS SSM: Stewardship of Patient Resources"

CAHPS Measure	All CAHPS Measures
1	"CAHPS for MIPS SSM: Getting Timely Care, Appointments and Information"
2	"CAHPS for MIPS SSM: How Well Providers Communicate"
3	"CAHPS for MIPS SSM: Patient's Rating of Provider"
4	"CAHPS for MIPS SSM: Health Promotion and Education"
5	"CAHPS for MIPS SSM: Care Coordination"
6	"CAHPS for MIPS SSM: Courteous and Helpful Office Staff"
7	"CAHPS for MIPS SSM: Stewardship of Patient Resources"
8	"CAHPS for MIPS SSM: Between Visit Communication"

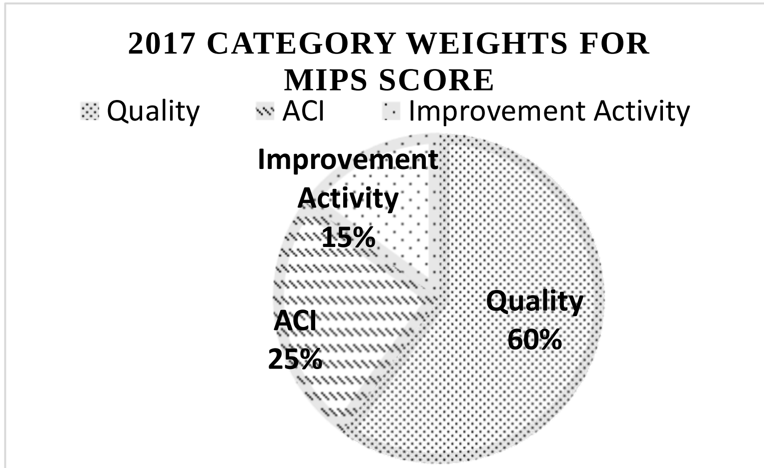
Appendix 2

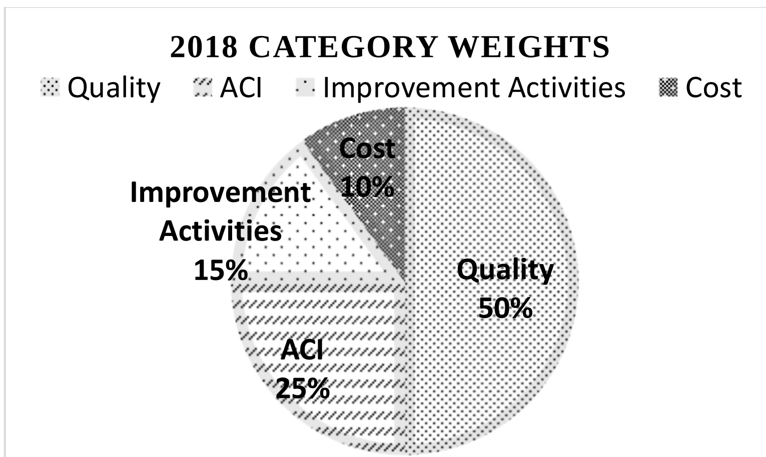
Details of MIPS

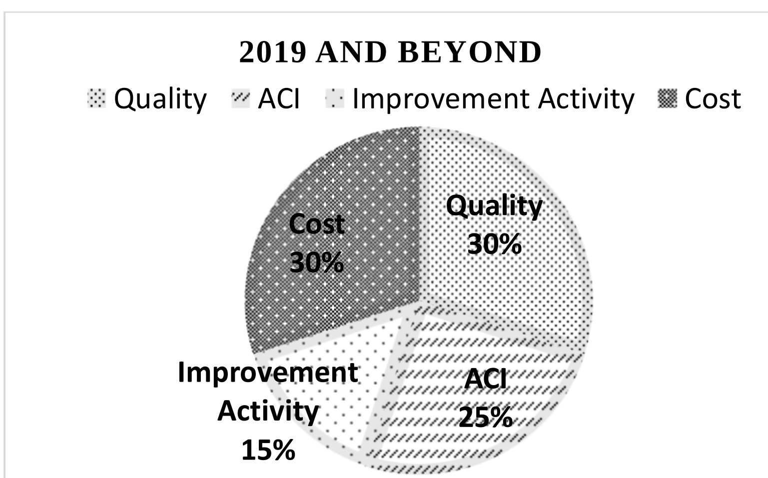
The Medicare Access and CHIP Reauthorization Act of 2015 (MACRA) established a quality payment incentive program, termed the Quality Payment Program (QPP). This payment program rewards providers for better value care and patient outcomes to incentivize a shift toward value-based payment systems and away from volume-based fee-for-service payment models within the healthcare industry. Doctors and other clinicians may receive higher Medicare payments based on their performance and engagement in select activities as a result of this program. The QPP has two options from which to choose: the Merit-based Incentive Payment System (MIPS) and the Advanced Alternative Payment Model (Advanced APMs).

Under MIPS, the final Medicare payment amount is calculated based on the provider's performance score on four categories: quality, improvement activities, advancing care information, and cost reduction. Each category is weighted differently when determining the final score and payment amount; the category weights are set to adjust beyond the transition period (see pie charts below).

Figures 12, 13, 14







The Quality category under MIPS, which is under Medicare's Quality Payment Program, asks that providers select from a list of 271 activities or "measures," that coincide with health care quality goals. The quality goals for care are the following: effective, safe, efficient, patient-centered, equitable, and timely. Quality measures track either the process or the outcome of care, and some can be classified as "high-priority." The Center for Medicare and Medicaid Services (CMS) gives the provider a quality performance score based on the measure data chosen and submitted by the provider. Of the four performance categories, quality is the most heavily weighted (at 60% of the final MIPS score in year 2017).

MIPS quality score replaces the Physician Quality Reporting System (PQRS), and the Physician Value-based Payment Modifier (VM). To transition in 2017 from the outdated system, individual reporting providers can choose one of six data submission mechanisms: Qualified Clinical Data Registry (QCDR), Qualified Registry, Electronic Health Record (EHR), claims, CMS website, and CAHPS for MIPS survey vendors. Also, clinicians decide whether to report data for only a small window of time called a "test period." If clinicians participate beyond the test period, they must submit data on at least six measures. Clinicians have the freedom to select the measures on which they would like to be scored. There are up to 127 measures from which to choose in 2017.

Furthermore, if eligible clinicians do not submit at least six measures or any outcome or high priority measure, CMS will conduct an Eligibility Measure Applicability (EMA) process*. If the EMA process finds that additional clinically related measures could have been submitted and were withheld, the provider's Quality performance category score will change. This insight is relevant to my study as it disincentivizes providers to “cherry-pick” their quality metrics, at least for some portion of the eligible clinician population. Perhaps avoidance of certain submission mechanisms could result from this policy or perhaps, in the future, the EMA process will apply to all submission mechanisms in order to capture all intentional exclusion of relevant measures.

**specific to the data submission mechanism (and only applies to claims and Qualified Registry submission mechanisms). In CY 2017, under the “pick your pace option,” no negative payment adjustment will occur as a result of the EMA (Writers, 2019).*

**EMA replaces the Measure-Applicability Validation (MAV) process used for the Physicians Quality Reporting System (PQRS).*

Scoring the Quality Category (for CY 2017)

Performance on quality measures is classified into 10 “deciles,” which are stratified using the national performance benchmarks from two years prior. For example, the quality data benchmarks for CY2017 are based on data reported through PQRS in CY2015. The benchmarks for CAHPS for MIPS survey are based on data from two past surveys: 2015 CAHPS for PQRS and the 2015 CAHPS for ACOs. Each decile then has a value between one and ten. Unless specified, “one” represents the lowest quality value. A provider’s performance on a quality measure is compared to the performance levels in the national deciles. Points (one through ten) are awarded to a provider based on the decile range that his or her performance level matches.

Each quality measure is converted into a ten-point scoring system. Depending on submission choice, a clinician can earn up to 60 or 70 points. For the transition year, clinicians will automatically receive a minimum of three points for completing and submitting at least one quality measure. The total quality performance score equals the summation of the points earned on the required six quality measures and any bonus points earned, divided by the maximum number of points.