NBER WORKING PAPER SERIES

GAMING AND EFFORT IN PERFORMANCE PAY

Luca Bertuzzi
Paul J. Eliason
Benjamin Heebsh
Riley J. League
Ryan C. McDevitt
James W. Roberts

Working Paper 31353
http://www.nber.org/papers/w31353

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2023

We thank nephrologists Matthew Ellis, Ruediger Lehrich, John Middleton, and Myles Wolf for providing valuable industry and medical insights. We gratefully acknowledge research support provided by the Social Sciences Research Institute and the National Science Foundation (SES 1850736 and 1851615). We thank Jeffrey Hill, Alexander Marsh, and Gabor Palinko for their excellent research assistance. The data reported here have been supplied by the United States Renal Data System (USRDS). The interpretation and reporting of these data are the responsibility of the authors and in no way should be seen as an official policy or interpretation of the U.S. government, nor does the content necessarily reflect the views of the National Bureau of Economic Research.

Gaming and Effort in Performance Pay
Luca Bertuzzi, Paul J. Eliason, Benjamin Heebsh, Riley J. League, Ryan C. McDevitt, and James W. Roberts
NBER Working Paper No. 31353
June 2023
JEL No. I11,I13,L10,L15,L2,L50

ABSTRACT

Health insurers often tie payments to providers' quality of care. Although payers do this to elicit more effort from providers, some providers may game the system by avoiding patients who would cause their quality scores to fall. We use annual variation in the criteria for Medicare's Quality Incentive Program in dialysis to distinguish strategic patient dropping from higher-quality care. Patients who would reduce their facilities' scores are 14.3–71.5% more likely to switch facilities, often to ones that suggest the move was involuntary, while under certain conditions facilities exert more effort to improve their scores by providing better care.

Luca Bertuzzi
Duke University
luca.bertuzzi@duke.edu

Paul J. Eliason
Department of Economics
Brigham Young University
2147 WVB
Provo, UT 84602
and NBER
paul.eliason@byu.edu

Benjamin Heebsh
Federal Trade Commission
bheebsh.research@gmail.com

Riley J. League
Duke University
Department of Economics
213 Social Sciences Building
Durham, NC 27708
riley.league@duke.edu

Ryan C. McDevitt
The Fuqua School of Business
Duke University
100 Fuqua Drive
Durham, NC 27708
and NBER
ryan.mcdevitt@duke.edu

James W. Roberts
Duke University
Department of Economics
213 Social Sciences Building
Durham, NC 27708
and NBER
j.roberts@duke.edu

1 Introduction

Health insurers routinely tie payments to providers' quality of care. Although designed to give providers an incentive to exert more effort in caring for patients, pay-for-performance reimbursement schemes may spur providers to avoid treating patients who would hurt their overall performance scores. Such gaming of quality benchmarks not only undermines the integrity of pay-for-performance systems, it also directly harms patients by disrupting their care and forcing them to seek out less-preferred providers. In this paper, we study Medicare's End-Stage Renal Disease Quality Incentive Program (ESRD QIP) in dialysis to show directly how pay-for-performance models induce both effort and gaming among providers.

Medicare started the QIP in 2012 following its move to a prospective payment system for dialysis. Before then, Medicare had used a hybrid model that paid facilities a fixed reimbursement for each dialysis session and a fee-for-service reimbursement for any injectable drugs administered during treatment, with excessive doses of separately billable drugs such as EPOGEN prompting calls for reform (Eliason et al., 2023). To address concerns that facilities might undertreat patients once the payment reform made injectable drugs a marginal cost rather than source of profit, Medicare began assigning each facility an annual quality score based on a set of standardized metrics, such as hemoglobin levels and dialysis adequacy, while cutting reimbursements for facilities that fall below certain thresholds by up to 2%.

Because Medicare frequently updates the metrics used to evaluate facilities, the QIP offers an ideal setting to identify how providers respond to performance pay: the annual changes generate substantial variation in patients' quality scores, and patients are much more likely to switch facilities in the years in which their characteristics would trigger a penalty, a correlation consistent with claims that facilities involuntarily discharge less-profitable patients (Fields, 2010). Along some dimensions, however, we find that facilities exert more effort to improve their quality of care during the years in which poor performance on that specific measure would be penalized. In short, performance pay leads to both real improvements in quality as well as strategic patient dropping to game Medicare's quality scores.

Two performance measures in particular illustrate this phenomenon. Dialysis facilities must regularly track the amount of waste and toxins they clear from a patient's blood, with dialysis adequacy measured by either the urea reduction ratio (URR) or Kt/V. These two measures are closely, but not perfectly, related: patients with an identical URR might end up having very different measures of Kt/V, primarily due to differences in body weight. Initially, the QIP used URR as its performance measure for dialysis adequacy, but then switched to Kt/V in the program's fourth year and set off facilities' strategic responses to performance pay. When the QIP penalized poor URR scores, patients with bad URR and good Kt/V were more likely to switch facilities than those with good URR and bad Kt/V; when the QIP penalized poor Kt/V instead, the pattern flipped.

Facilities' incentives for dropping penalty-inducing patients under the QIP are clear-cut. A facility with the average load of 50 eligible patients and at the bottom decile of QIP scores could move from the worst possible score to the best by dropping just three patients. Similar to the example of dialysis adequacy, we use Medicare claims data to detect whether facilities engage in such strategic behavior across all QIP criteria by calculating a penalty score for each patient based on how much their diagnostic measures would reduce their facility's QIP score in a given year. In our most conservative specification the controls for patients' contemporary health characteristics, we find that a patient who does not satisfy any of Medicare's QIP parameters is 14.3% more likely to switch facilities than a patient who satisfies them all. We also find that penalty-inducing patients are more likely to switch facilities after being hospitalized — a result consistent with facilities selectively refusing to take back patients who might harm their QIP scores — and that these penalty-inducing patients are not more likely to switch to more-convenient or higher-quality facilities, suggesting their move was unlikely to be voluntary.

Along with strategically dropping patients, facilities can also improve their QIP scores by exerting more effort to provide better care. To improve adequacy scores, for example, a facility could increase the amount of time their patients spend being dialyzed, although doing

so comes with the opportunity cost of not using the dialysis station to treat another patient; the facility must then weigh the financial penalty of the QIP against the forgone payment from not treating an additional patient. For the QIP measures where we have data directly related to effort, we find facilities exert more effort to improve their quality of care when the incentives from the QIP do not conflict with other profit-maximizing activities. During the years in which the QIP penalized facilities for having patients with low hemoglobin levels, for instance, facilities administered higher doses of the fee-for-service drugs that stimulate red blood cell production, allowing them to avoid the reimbursement cut from falling short of the QIP's benchmarks while at the same time earning higher profits from the separately billable drugs. Similarly, when the QIP penalized hypercalcemia, facilities exerted more effort to reduce calcium levels by prescribing more cinacalcet, a drug covered under Part D for which the patient, rather than the facility, bears the cost. When the QIP's incentives conflict with facilities' other profit motives, like with run times for dialysis adequacy or anti-anemia drugs for transfusions, we do not find the same unambiguous increase in effort.

The growing use of pay-for-performance schemes in health care has prompted a recent literature studying their effects. Norton et al. (2018), for instance, find evidence that the Hospital Value-Based Purchasing Program spurred hospitals to improve their performance over time in the areas where they have the highest marginal incentives to do so, while Li and Norton (2019) find that measures in the Home Health Value-Based Purchasing Program improved by approximately 1% as agencies manipulated their coding of patients in ways that inflated their performance. Also related to gaming, Lisi et al. (2020) suggest that pay-for-performance schemes that reward low mortality or readmission rates can either weaken or strengthen a hospital's incentive to provide high-quality care, depending on selection bias.

Perhaps most closely related to our paper, Gupta (2021) uses Medicare claims data to study the Hospital Readmissions Reduction Program (HRRP), which penalizes hospitals with high readmission rates. Gupta finds that hospitals' responses to the penalty account for two-thirds of the decrease in readmissions over the study period in addition to a decrease

in mortality for heart attack patients. Half of these gains come from quality improvements and the remainder from differences in the mix of returning patients.

Our paper builds on and complements these existing studies in at least two important ways. First, the QIP comprises a rich set of measures that varies over time, allowing us to use Medicare's changes to the measures that make up a facility's QIP score to separately identify how facilities treat patients when their specific characteristics make them more likely to trigger a penalty. Second, the detailed claims data available for Medicare's dialysis patients, along with the chronic nature of the condition, allow us to observe the direct ways in which facilities change their effort for the same patient in response to performance pay as opposed to inferring them from aggregate spending measures.

Our paper also contributes to the growing literature on cream skimming and patient dropping in health care more generally. Policymakers typically view risk-adjusting reimbursements as a way to mitigate incentives for patient selection, but providers may game the payment scheme by choosing patients with poorly calibrated risk scores that make them more profitable to treat (Geruso et al., 2019; Brown et al., 2014). Alternatively, high-powered bonuses to decrease total treatment costs can prompt providers to select healthier patients and identify low-cost patients within risk-adjustment groups (Alexander, 2020), and providers can game pricing changes for targeted patient categories (Dafny, 2005; Eliason et al., 2018). Similar behavior has been found when providers face capacity constraints (Yang et al., 2020; Hackmann et al., 2020; Gandhi, 2019) and using randomized field experiments (Werbeck et al., 2021).

Finally, our paper contributes to the literature focused specifically on the impact of the QIP (Ajmal et al., 2020; Fink et al., 2002; Ajmal et al., 2019; Saunders et al., 2017), which has largely criticized the program for excluding measures of patient satisfaction. Along with the contemporaneous work of Kepler et al. (2022) on facilities' location decisions, we believe our paper is the first to directly investigate the impact of the QIP on the strategic behavior of facilities, including patient dropping and investments in quality to improve health outcomes.

2 Institutional Details and Data

2.1 Kidney Failure and the Dialysis Industry

Kidneys filter wastes and toxins out of the blood and stimulate red blood cell production. For patients experiencing end-stage renal disease, the kidneys no longer adequately perform these functions, which necessitates either a kidney transplant or dialysis. The most common form of dialysis, hemodialysis, uses a machine to mechanically filter wastes and toxins from a patient's blood, either at the patient's home or at a medical facility. Nearly all ESRD patients also receive a cocktail of drugs to address common comorbid conditions, such as anemia and hypercalcemia.

Patients primarily receive dialysis at one of the more than 7,000 dedicated dialysis facilities across the country, where they typically go three times per week for treatment that lasts three to four hours each visit. These facilities are run by a mix of for-profit and non-profit firms, with the two largest for-profit chains, DaVita and Fresenius, controlling over 60% of facilities and earning 90% of the industry's revenue (United States Renal Data System, 2014; Baker, 2019; Eliason et al., 2020).

2.2 Medicare Payment Reform

Since 1972, Medicare has extended full benefits to all patients with ESRD, regardless of age, paying for outpatient dialysis and anemia treatment under Part B.[1] From the early 1980s to 2010, Medicare paid providers a flat rate for each dialysis session and a fee-for-service reimbursement for any injectable drugs administered during treatment. Concerns that the distortionary incentives from fee-for-service reimbursements resulted in excessive costs for Medicare and harm to patients motivated policy makers to include ESRD payment reform as part of the Medicare Improvements for Patients and Providers Act (MIPPA) in

[1]Those enrolled in an employer group health plan when diagnosed with ESRD retain their commercial insurance as a primary payer for 33 months, during which time Medicare acts as a secondary payer (League et al., 2022).

2008, which mandated the bundling of dialysis services and all injectable drugs and biologics used in the treatment of ESRD into a single prospective payment, starting in 2011.[2]

To offset the incentives for providers to reduce costs by providing lower-quality care, MIPPA also mandated the development of the QIP, which links payments to performance by reducing reimbursements to low-performing facilities by up to 2%. To determine penalties, Medicare constructs a score based on three time periods: a historical "comparison year" that provides a baseline for measuring quality improvements, a "performance year" in which the facility's quality is evaluated, and a "payment year" two years after the performance period when a low-performing facility receives its penalty.[3]

To calculate a facility's total performance score, Medicare compares the share of patients at the facility who satisfy a given standard in the performance year to either a national facility-level average or the facility's own past performance in the comparison year, with the facility receiving the greater of the two scores; the facility's scores on each measure are then aggregated based on their respective weights in the QIP that year. Facilities with a total performance score just below the administratively set threshold have their payments reduced by 0.5% and those far below by 2%, with facilities in between facing 1% or 1.5% reductions.[4]

Although the basic scoring mechanism has remained constant since the introduction of the QIP, the actual performance measures used to construct the scores have changed over time along with the weights placed on them in calculating the final score, as shown in Figure 1a. In the first performance year, Medicare used three quality measures related to clinical outcomes: the percentage of patients with average hemoglobin values above 12, the percentage with average hemoglobin values below 10, and the percentage of patients

[2]Eliason et al. (2023) study how the payment reform affected the most common of these drugs, EPOGEN.

[3]The first two years of the program use a comparison period for the national benchmark two years before the performance year, whereas each facility is compared to its own performance in 2007 for both years. In the third year of the QIP, the national benchmark and each facility's own performance benchmarks were constructed using performance from July 2010 to June 2011. For all subsequent years, the national comparison period is the year before the self-comparison period, which is the year before the performance period. In every year, new QIP measures are finalized with no more than two months remaining in the comparison period. Online Appendix Figure A2 provides more details on these timelines.

[4]Online Appendix A provides more details on the QIP scoring system.

with a median urea reduction ratio (URR) below 65%. Medicare then stopped penalizing hemoglobin levels below 10 after the first year following a black box warning from the FDA that changed the standard of care for anemic dialysis patients (Eliason et al., 2023), and over time Medicare has updated its measures and weights in response to new medical guidelines and the desire to provide more transparency for patients.

As will be important for our identification strategy below, Figure 1b shows how the continual update of QIP rules generates substantial variation in the likelihood a patient would trigger a penalty for his or her facility. Because the new rules are typically proposed and finalized very close to the performance period — or even during the performance period for the first years of QIP — many patients suddenly cause their facilities' scores to fall solely due to exogenous changes in what Medicare evaluates rather than due to changes in their underlying attributes, whereas others suddenly become compliant on all measures and no longer trigger a penalty. The elevated rate of switching for penalty-inducing patients in Figure 1c would then be consistent with facilities strategically dropping patients in response to the QIP.

2.3 Data

The main data for our analysis come from the U.S. Renal Data System, which collects and manages data from a variety of sources relevant to ESRD patients and health care providers.[5] Included in these data are Medicare claims, treatment histories, patient attributes, and annual facility surveys. In addition, CMS Form 2728, known as the Medical Evidence Form, provides rich information on the health and clinical attributes of patients when they begin dialysis. Our sample includes 1,018,413 patients and 7,770 facilities, with summary statistics appearing in Online Appendix B.

To identify strategic patient dropping, we begin by classifying patients who switch facilities. On average, patients rarely do so — only about 0.8% switch in a given month —

[5]For a more thorough description of USRDS, please see the *Researcher's Guide to the USRDS System* at USRDS.org (United States Renal Data System, 2020).

primarily because they must receive dialysis three times per week and strongly prefer going to a facility close to their homes. To avoid including temporary switches in our analysis (e.g., if a patient receives dialysis while on vacation), we only use switches that occur within a patient's hospital referral region for which a patient does not return to his or her original facility within the following six months. Although Medicare does not release data on the reason a patient switches to a new facility, we use a proxy for involuntary switches based on whether a patient's new facility is farther away or has worse outcomes, as such moves would seem to be strictly dominated based on the assumption that patients prefer closer, higher-quality facilities (Eliason, 2022).

We use the QIP scoring rules to determine whether a patient increases the likelihood that a facility receives a payment reduction in the relevant performance year. Specifically, we construct a penalty score for each patient based on the number of measures for which a patient falls short of the QIP standard as well as the relative weights of those measures. To do this, we sum a series of indicator functions for whether the patient fell short of each QIP standard in effect at time t, weighted by the respective QIP weights and scaled to span zero to one. Once again, Figure 1b shows how the share of patients who failed to meet at least one benchmark — and who therefore have a penalty score greater than zero — has evolved over the years in our data, with the large changes in the share of penalty-inducing patients from year to year demonstrating how changes in QIP performance measures shift patients in and out of penalty-inducing status. Underpinning these annual changes, Table 1 provides further summary statistics for the evolution of each QIP measure, where the bold green cells highlight the years in which that measure appeared in the QIP. In general, dialysis facilities' quality of care has improved over time, with the share of patients failing to meet the QIP standards for most measures falling after being included in the QIP calculation. Table 1 also suggests that all measures have a material impact on facilities' scores, as opposed to a single measure being the sole determinant of patients inducing a penalty in a given year.

3 Empirical Analysis

We use Medicare's annual changes in QIP criteria to estimate the causal effect of pay-for-performance measures on the likelihood that patients switch facilities and facilities exert more effort. As shown in Figure 1c, penalty-inducing patients are more likely to switch facilities relative to those who do not induce penalties, which motivates our main specification,

$$(1) \qquad Y_{ijt} = \beta I_{\text{Penalty},t}\left[Measures_{it}\right] + \alpha_1 Measures_{it} + \alpha_2 Measures_i^{Base} + \alpha_3 X_{ijt} + \varepsilon_{ijt},$$

where Y_{ijt} is the outcome of interest for patient i at facility j in month t, and X_{ijt} includes a host of facility and patient controls in addition to facility and year-month fixed effects.

The key outcomes we study are the probability that a patient switches dialysis facilities, the characteristics of that switch, and the amount of effort exerted by facilities to treat the patient. To capture how pay-for-performance affects facilities' choices, we include $I_{Penalty,t}[Measures_{it}]$, a function that maps all of the patient's performance measures to their penalty score based on the QIP criteria in effect at time t. To account for the possibility that switching could be directly related to the factors included in the QIP performance measures, we control for $Measures_{it}$, the patient's contemporary values on all performance measures that ever enter the QIP, as well as $Measures_i^{Base}$, the patient's average values on each performance measure during his or her first six months on dialysis.

The parameter of interest, β, gives the differential outcomes for patients who potentially induce penalties for their facilities under the QIP. The key challenge in identifying β is that penalty-inducing patients may have different facility-switching behavior or receive different care than other patients due to differences in their underlying health rather than differences in facilities' financial incentives to treat them. We overcome these challenges by (i) including a rich set of patient and facility characteristics that allow us to compare observably similar patients, (ii) including time fixed effects that allow us to compare the outcomes for patients who have different penalty scores in the same month, uncontaminated by universal changes

in outcomes over time, and (iii) including facility fixed effects that allow us to capture any time-invariant, unobserved factors that might make a facility more predisposed to patients leaving, such as operating in an area with more vulnerable patients, or exert different levels of effort, depending on the outcome considered.

At the same time, penalty-inducing patients may have different outcomes due to unobserved differences not captured in X_{it}. For example, if patients who are more prone to disregard medical advice are also more likely to switch facilities voluntarily, then these patients may induce a correlation between negative health outcomes that trigger a penalty and their decision to switch facilities for reasons unrelated to performance pay. The detailed claims data, repeated patient encounters, and timing of QIP criteria changes allow us to overcome these threats to identification as well.

First, we control for patients' baseline characteristics on all QIP performance measures, which allows us to compare patients who begin dialysis with the same health status but who may have different penalty scores depending on which measures appear in the QIP in different years. Patients who start dialysis with the same potential to induce penalties — that is, patients who initiate dialysis with the same baseline characteristics but before facilities' strategic behavior could affect their health outcomes — will differ in their actual penalty score as their health statuses evolve over time and different measures enter and exit the QIP. For example, patients who have low hemoglobin levels when they begin dialysis will differ in their penalty score depending on whether they started in 2010–2011, when low hemoglobin was included in the QIP, or after 2011, when that measure was no longer one of the quality standards.

Even more conservatively, we can also include $Measures_{it}$ in the regression to control for a patient's contemporary health characteristics that, depending on the QIP criteria in effect at the time, determine their penalty scores. Directly controlling for these characteristics in the regression means that we identify the impact of the penalty score using only changes in which criteria are being evaluated by Medicare in that year rather than by comparing

patients with different health statuses. This strategy is similar in spirit to a difference-in-differences analysis that compares the change in switching behavior after a QIP measure is introduced across patients who do or do not satisfy the standard. Including contemporary measures in this way may understate the effect of strategic patient dropping, however, as any extra effort exerted by facilities in response to the QIP may improve a patient's outcomes to the point where it is no longer necessary to cut ties with them.

Returning to the example of dialysis adequacy, Figure 1d clearly demonstrates our identification strategy. Prior to 2013, Medicare used URR to calculate facilities' QIP scores for dialysis adequacy, and patients with URR values below the threshold of 65% during this time were more likely to switch facilities than those who achieved the QIP standard. Medicare then replaced URR with Kt/V in 2013, and despite clinicians using both measures to assess the adequacy of a patient's treatment, the slight differences between them mean that some patients meet the QIP threshold for one but not the other. Following the QIP update, the probability of switching facilities increased for patients who met the old standard but failed to achieve the new one (i.e., they had good URR but low Kt/V), whereas patients who suddenly met the new standard but who would have failed before (i.e., they had good Kt/V but low URR) suddenly began switching at a lower rate.

Moving beyond our case study of dialysis adequacy, Table 2 presents the results of estimating equation (1) using the full set of QIP measures. We find across all specifications that penalty-inducing patients are more likely to switch facilities. In column (1), patients with the highest penalty scores are 71.5% more likely to switch facilities than patients with the lowest penalty scores, while a more conservative specification that includes contemporary values for QIP measures in column (2) shows an elevated switching rate of 14.3%.[6] Alternatively, a one standard deviation increase in a patient's penalty score is associated

[6]Including contemporary measures in column (2) dampens the effect of penalty scores on switching for two reasons. First, by adding contemporary QIP measures among the controls, we identify the impact of being penalty-inducing using only changes in which criteria are being evaluated rather than by comparing patients with different contemporary health statuses. Second, the observations we drop due to missing contemporary values have higher penalty scores and worse baseline QIP values for most measures.

with a 10.0% higher probability of switching in column (1) and a 2.0% higher probability in column (2). Restricting the sample to hospitalized patients in columns (3) and (4), we again find that patients with the highest penalty scores are more likely to switch facilities, which ranges from 19.4–51.9% depending on the specification. As hospitalizations represent an opportunity for facilities to cut ties with a patient who might harm their QIP score by selectively refusing to take back those with unfavorable characteristics, this result is again consistent with strategic patient dropping.

Our results likely understate strategic patient dropping in two important ways. First, we include all facilities in our analysis even though many are not at risk of being penalized (i.e., a marginal patient would not put them at risk of triggering a payment reduction). Second, facilities may exert effort to encourage certain penalty-inducing patients to leave, but we only observe when this effort results in a successful dropping.

To determine whether a switch is unlikely to be voluntary, we compare the mortality, hospitalization, and infection rates of the facility a patient leaves with his or her new facility, as well as whether the new facility is closer to the patient's home, with the intuition that a patient would not willingly move to a lower-quality or farther-away facility. Table 3 presents the results of estimating equation (1) restricted to a sample consisting solely of patient switches and replacing Y_{ijt} with the difference in the outcome variable between the new facility and the original one. Compared to an average annual mortality rate of 8%, column (1) suggests that patients with the highest penalty scores move to facilities with a 5.9% higher mortality rate than patients with the lowest scores. Although the difference vanishes when we add contemporary values in column (2), the missing values may bias us against finding a positive effect and the coefficient is not statistically significant — that is, we can still rule out the alternative explanation that patients with higher penalty scores are switching to receive care at a better facility. We find similar patterns for our other two measures of quality, infection and hospitalization rates, while column (7) shows that patients with higher penalty scores travel an additional mile after switching compared to a baseline of 11 miles.

Finally, Table 4 shows the extent to which facilities exert more effort to avoid penalties by providing better care to patients, where the dependent variable of each column is a specific type of effort or activity that can be undertaken by the facility to help satisfy an associated QIP measure, and the explanatory variable $\mathbb{1}\{\text{Penalty-Inducing}\}$ becomes one if the associated QIP measure is not satisfied. To circumvent issues that would arise from simultaneous efforts to improve care and drop patients, this specification uses October–December of the previous year to construct a patient's QIP status for each measure. For the low hemoglobin criteria, for example, $\mathbb{1}\{\text{Penalty-Inducing}\}$ is equal to one in 2010 if a given patient has a low hemoglobin reading from October 2009 through December 2009, and zero otherwise, while the variable $\mathbb{1}\{\text{Post}\}$ turns on to one after the measure has been added to the QIP.

We find effort increases for those measures for which avoiding QIP penalties aligns with other profit-maximizing activities: EPOGEN and iron doses increase for patients with low hemoglobin levels at a time when fee-for-service reimbursements made these higher doses profitable for facilities to administer; EPOGEN and iron doses decrease for patients with high hemoglobin levels at a time when prospective payments made these lower doses a cost savings for facilities; and cinacalcet doses increase for patients with high calcium readings at a time when patients bear their expense under Part D rather than the facilities. In short, facilities exert more effort to avoid penalties when they can do so in ways that do not otherwise reduce their profits.

When facilities face a tradeoff between avoiding QIP penalties and incurring additional costs, however, we do not find the same unambiguous increase in effort. For inadequate Kt/V scores, increasing treatment time decreases the likelihood of substandard Kt/V levels, but doing so means the facility cannot treat another patient during that time to receive another payment from Medicare. Reflecting this opportunity cost, treatment time does not increase after Medicare added Kt/V to the QIP criteria, with the switching trends in Figure 1d suggesting that facilities instead prefer to game the system by dropping patients rather than

incur additional costs to improve their quality. The decline in EPOGEN doses for penalty-inducing patients after transfusions enter the QIP similarly reflects this logic. That is, higher EPOGEN doses decrease the risk of patients requiring transfusions and triggering a penalty, but the QIP payment reduction is likely to be much less than the cost of administering such large doses across a wide swath of patients. Iron replacement drugs, on the other hand, are a cost-effective way to stimulate red blood cell production and avoid the penalties for transfusions under the QIP, and we find that these doses increase for transfusion patients when the QIP provides an incentive for facilities to do so.

4 Conclusion

Patients who would trigger a reimbursement penalty are much more likely to switch facilities after the introduction of pay-for-performance in dialysis. The switches appear to be involuntary, as penalty-inducing patients are more likely to move to worse facilities that are farther away from their homes.

One reason the QIP leads to gaming by facilities is misaligned financial incentives. Facilities can ensure adequate Kt/V by increasing treatment times, for instance, but doing so comes with the opportunity cost of treating fewer patients and receiving commensurately lower prospective payments. In such cases, facilities find it more profitable to drop patients who would trigger QIP penalties rather than invest more effort to improve their outcomes.

Our findings highlight the need for Medicare and other health insurers to account for providers' strategic behavior when designing pay-for-performance reimbursement schemes. When providers can manipulate their scores through strategically selecting patients, the disruption to patient care may outweigh the benefits that result from the increase in effort. Penalizing facilities for patients who switch facilities may be one possible remedy, as would a look-back period that assigns patients to all facilities that recently provided care to them.

References

F. Ajmal, J. Probst, J. Brooks, J. Hardin, and Z. Qureshi. Association between freestanding dialysis facility size and medicare quality incentive program performance scores. *American journal of nephrology*, 49(1):64–73, 2019.

F. Ajmal, J. C. Probst, J. M. Brooks, J. W. Hardin, Z. Qureshi, and T. H. Jafar. Freestanding dialysis facility quality incentive program scores and mortality among incident dialysis patients in the united states. *American Journal of Kidney Diseases*, 75(2):177–186, 2020.

D. Alexander. How do doctors respond to incentives? unintended consequences of paying doctors to reduce costs. *Journal of Political Economy*, 128(11):4046–4096, 2020.

S. Baker. The U.S. Health Care System is Full of Monopolies. *Axios*, 2019.

J. Brown, M. Duggan, I. Kuziemko, and W. Woolston. How does risk selection respond to risk adjustment? new evidence from the medicare advantage program. *American Economic Review*, 104(10):3335–64, 2014.

L. S. Dafny. How do hospitals respond to price changes? *American Economic Review*, 95 (5):1525–1547, 2005.

P. Eliason. Price Regulation and Market Structure: Evidence from the Dialysis Industry. Working Paper, 2022.

P. J. Eliason, P. L. Grieco, R. C. McDevitt, and J. W. Roberts. Strategic patient discharge: The case of long-term care hospitals. *American Economic Review*, 108(11):3232–65, 2018.

P. J. Eliason, B. Heebsh, R. C. McDevitt, and J. W. Roberts. How acquisitions affect firm behavior and performance: Evidence from the dialysis industry. *The Quarterly Journal of Economics*, 135(1):221–267, 2020.

P. J. Eliason, B. Heebsh, R. J. League, R. C. McDevitt, and J. W. Roberts. Measuring clinical effectiveness in practice: Evidence from EPO in dialysis. Technical report, 2023.

R. Fields. God help you. you're on dialysis. *The Atlantic*, 306(5):82–92, 2010.

J. C. Fink, M. Zhan, S. A. Blahut, M. Soucie, and W. M. McClellan. Measuring the efficacy of a quality improvement program in dialysis adequacy with changes in center effects. *Journal of the American Society of Nephrology*, 13(9):2338–2344, 2002.

A. Gandhi. Picking your patients: Selective admissions in the nursing home industry. *Available at SSRN 3613950*, 2019.

M. Geruso, T. Layton, and D. Prinz. Screening in contract design: Evidence from the aca health insurance exchanges. *American Economic Journal: Economic Policy*, 11(2):64–107, 2019.

A. Gupta. Impacts of performance pay for hospitals: The readmissions reduction program. *American Economic Review*, 111(4):1241–83, 2021.

M. B. Hackmann, R. V. Pohl, and N. R. Ziebarth. Patient versus provider incentives in long-term care. 2020.

J. D. Kepler, V. V. Nikolaev, N. Scott-Hearn, and C. R. Stewart. Quality transparency and healthcare competition. *Available at SSRN 3963418*, 2022.

R. J. League, P. Eliason, R. C. McDevitt, J. W. Roberts, and H. Wong. Variability in prices paid for hemodialysis by employer-sponsored insurance in the us from 2012 to 2019. *JAMA Network Open*, 5(2):e220562–e220562, 2022.

J. Li and E. C. Norton. Pay-for-performance and long-term care. In *Oxford Research Encyclopedia of Economics and Finance*. 2019.

D. Lisi, L. Siciliani, and O. R. Straume. Hospital competition under pay-for-performance: Quality, mortality, and readmissions. *Journal of Economics & Management Strategy*, 29 (2):289–314, 2020.

E. C. Norton, J. Li, A. Das, and L. M. Chen. Moneyball in medicare. *Journal of health economics*, 61:259–273, 2018.

M. R. Saunders, H. Lee, and M. H. Chin. Early winners and losers in dialysis center pay-for-performance. *BMC health services research*, 17(1):1–9, 2017.

United States Renal Data System. 2014 USRDS Annual Data Report: Epidemiology of Kidney Disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2014.

United States Renal Data System. USRDS 2020 Annual Data Report: Epidemiology of Kidney Disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2020.

A. Werbeck, A. Wübker, and N. R. Ziebarth. Cream skimming by health care providers and inequality in health care access: Evidence from a randomized field experiment. Technical report, National Bureau of Economic Research, 2021.

O. Yang, M. K. Chan, T. C. Cheng, and J. Yong. Cream skimming: Theory and evidence from hospital transfers and capacity utilization. *Journal of Economic Behavior & Organization*, 173:68–87, 2020.

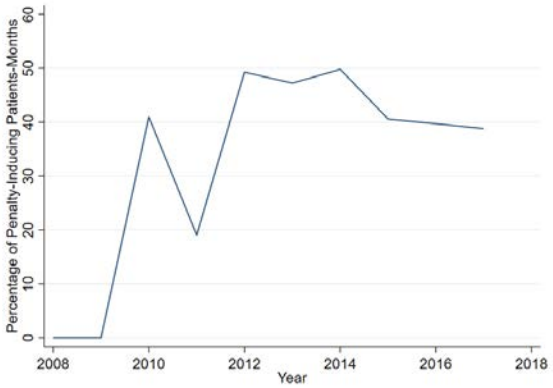
Tables & Figures

Figure 1: QIP Criteria and Switching

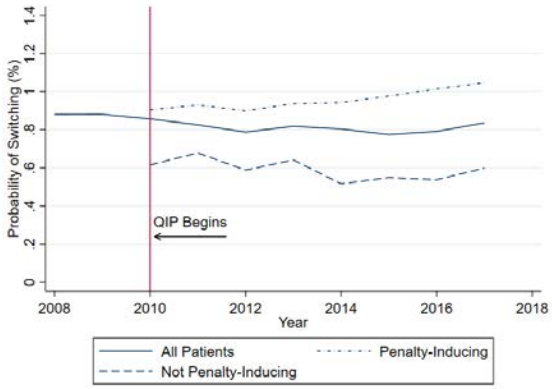
(a) QIP Criteria and Weights

		Performance Year							
Measure	Description	2010	2011	2012	2013	2014	2015	2016	2017
Low HGB	% of patients with HGB<10 g/dL	50%							
High HGB	% of patients with HGB>12 g/dL	25%	50%	33%	33%	21.4%			
Low URR	% of patients with a median URR below 65%	25%	50%	33%					
Bad Vasc. Access	% of patients using a catheter or graft for vascular access			33%	33%	21.4%	21.4%	14.7%	15.6%
Low Kt/V	% of patients with Kt/V<1.2				33%	21.4%	21.4%	14.7%	15.6%
Infection	% of patients with a positive blood culture					21.4%	21.4%	23.5%	15%
Transfusion	Ratio of observed and predicted transfusions							14.7%	15.6%
Readmission	Ratio of observed and predicted unplanned readmissions						21.4%	17.6%	22.6%
High Calcium	% of patients with 3-month mean of serum calcium>10.2 mg/dL					14.3%	14.3%	14.7%	15.6%

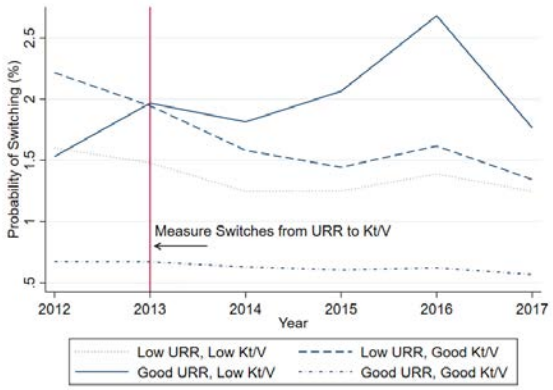
(b) Percentage of Penalty-Inducing Patients



(c) Switching by Penalty-Inducing Status



(d) Switching by URR and Kt/V Status



Notes: Subfigure (a) reports the set of QIP criteria active in each performance year and the weight each measure receives in the overall facility score. Subfigure (b) reports the percentage of penalty-inducing patient-months each year. Subfigure (c) reports the probability of switching facilities in a given patient-month by whether the patient fails to satisfy at least one QIP measure (i.e., is "penalty-inducing"). Subfigure (d) reports the probability of switching facilities by whether the patient satisfies the QIP standard for Kt/V and/or URR. Patient-months with a median URR below 65% have "Low URR," while patient-months with a final Kt/V reading below 1.2 have "Low Kt/V."

Table 1: Penalty-Inducing Status by QIP Measure

	Penalty-Inducing	Penalty Score	Low HGB	High HGB	Low URR	Bad Vasc. Access	Low Kt/V	Infection	Transfusion	Readmission	High Calcium
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
2008	0.0	0.0	8.1	32.7	8.8			2.1	0.6	0.5	
2009	0.0	0.0	8.8	27.4	7.8			2.0	0.6	0.5	
2010	***41.0***	***0.12***	***9.5***	***22.2***	***7.6***	43.4		2.0	0.6	0.5	
2011	***19.1***	***0.08***	13.6	***10.7***	***6.0***	41.3		1.9	0.3	0.5	
2012	***49.3***	***0.17***	17.4	***6.1***	***3.5***	***38.8***	3.5	1.8	0.4	0.4	4.1
2013	***47.2***	***0.15***	18.7	***4.8***	3.5	***36.5***	***3.1***	1.8	1.1	0.4	3.4
2014	***49.8***	***0.10***	19.6	***4.3***	3.5	***35.2***	***3.2***	***1.8***	1.0	0.3	***3.4***
2015	***40.6***	***0.09***	15.8	3.5	3.4	***35.1***	***3.0***	***1.8***	0.9	***0.3***	***2.5***
2016	***39.7***	***0.06***	10.1	2.2	3.0	***34.9***	***2.6***	***1.8***	***0.7***	***0.2***	***1.6***
2017	***38.8***	***0.07***	10.7	2.5	2.8	***34.7***	***2.1***	***1.8***	***0.6***	***0.2***	***1.4***

Notes: Column (1) reports the percentage of patient-months in which the patient does not meet at least one QIP standard in effect at that time. Column (2) reports the average weighted penalty score, which ranges from 0 to 1. Column (3) reports the share of patient-months with hemoglobin below 10.0 g/dL. Column (4) reports the share of patient-months with hemoglobin above 12.0 g/dL. Column (5) reports the share of patient-months with a median urea reduction ratio below 65%. Column (6) reports the share of patient-months with a catheter or graft. Column (7) reports the share of patient-months with Kt/V below 1.2. Column (8) reports the share of patient-months hospitalized for septicemia. Column (9) reports the share of patient-months in which the patient receives at least one blood transfusion. Column (10) reports the share of patient-months with an unplanned hospital readmission. Column (11) reports the share of patient-months in which the patient has calcium levels above 10.2 mg/dL. Cells reporting data for years in which each measure is used to determine facilities' QIP scores are bolded and italicized in green font. The sample consists of in-center hemodialysis patients between the ages of 18 and 100 with Medicare as their primary payer. Columns (1)–(5) and (8)–(10) include observations from January 2008 to December 2017. Column (6) includes observations from July 2010, the earliest date information on vascular access is available, to December 2017. Column (7) includes observations from January 2012, the earliest date information on Kt/V is available, to December 2017. Column (11) includes observations from May 2012, the earliest date information on calcium is available, to December 2017.

Table 2: Switching by Penalty Score

	Switch Facilities		Switch Facilities After Hospitalization	
	(1)	(2)	(3)	(4)
Penalty Score	0.00389***	0.000699**	0.00529***	0.00148
	(0.000168)	(0.000244)	(0.000689)	(0.000953)
Patient Controls	Yes	Yes	Yes	Yes
Facility FEs	Yes	Yes	Yes	Yes
Month-Year FEs	Yes	Yes	Yes	Yes
Base Measures	Yes	Yes	Yes	Yes
Contemporary Measures	No	Yes	No	Yes
R^2	0.00496	0.00483	0.0157	0.0137
Observations	16,113,581	15,256,031	1,153,032	1,050,044
Mean Dep. Var.	0.00544	0.00490	0.0102	0.00762

Notes: OLS estimates of equation (1). The dependent variable is an indicator variable that takes value 1 if a patient switches facilities in a given month in columns (1)–(2), and whether the switch occurs after a hospitalization in columns (3)–(4). The primary explanatory variable of interest is a patient's contemporary penalty score. An observation is a patient-month. Sample consists of observations from January 2008 to December 2017 for in-center hemodialysis patients between the ages of 18 and 100 with Medicare as their primary payer. The first six months of observations for each patients are removed from the sample. Patient controls include indicator variables for incident comorbidities and characteristics reported on medical evidence forms, including diabetes, hypertension, BMI bin, GFR bin, HGB bin, income bin, high albumin, cancer, drug use, alcoholism, smoking behavior, necessity of assistance, COPD, ASHD, PVD, ischemic heart disease, and congestive heart disease, along with patient race, gender, and cubic functions of age and dialysis tenure. Facility controls include age, HRR, whether the facility is freestanding or hospital-based, and chain ownership status. BMI bin, GFR bin, HGB bin and income bin represent respectively the BMI quintile, GFR quintile, HGB quintile, and the income quintile of the ZIP code of residence for each patient. Additional controls include year-month fixed effects, facility fixed effects, and measures of the patient's health status relative to each measure ever used in the QIP during the patient's first six months in the data. Finally, where indicated, controls also include the patient's contemporary health status relative to each measure ever used in the QIP. More specifically, these controls take the form of indicators for each specific measure which take value 1 if the patient has readings which do not satisfy the QIP rules in the current period. Standard errors clustered by patient are in parentheses. *, ** and *** indicate significance at the 5%, 1% and 0.1%, respectively. Contemporary "Bad Vasc. Access," "Low Kt/V," and "High Calcium" are imputed to equal 0 prior to July 2010, January 2012 and May 2012, respectively. The regressions include indicator variables for the periods in which each of these measures is missing.

Table 3: Switching to Less-Desirable Facility by Penalty Score

	Δ Mortality Rate		Δ Hospitalization Rate		Δ Infection Rate		Δ Distance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Penalty Score	0.00473**	-0.00461	0.00342	-0.00558	0.00602**	-0.00419	1.227***	0.114
	(0.00155)	(0.00253)	(0.00346)	(0.00578)	(0.00202)	(0.00324)	(0.371)	(0.601)
Patient Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Facility FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month-Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Base Measures	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contemporary Measures	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.289	0.300	0.340	0.357	0.301	0.317	0.180	0.193
Observations	68,213	56,920	68,213	56,920	68,213	56,920	78,070	65,871
Mean Dep. Var.	0.000950	0.000300	-0.0125	-0.0130	-0.000647	-0.00151	0.315	0.227

Notes: OLS estimates where the dependent variable is the difference between the patient's original facility and new one for mortality rate in columns (1)–(2), hospitalization rate in columns (3)–(4), infection rate in columns (5)–(6), and distance from the patient's home in columns (7)–(8). An observation is a patient-month. Sample consists of observations from January 2010 to December 2017 for in-center hemodialysis patients between the ages of 18 and 100 with Medicare as their primary payer, restricted further to patient-months in which the patient switches facilities in the next month. The first six months of observations for each patients are removed from the sample. Patient controls include indicator variables for incident comorbidities and characteristics reported on medical evidence forms, including diabetes, hypertension, BMI bin, GFR bin, HGB bin, income bin, high albumin, cancer, drug use, alcoholism, smoking behavior, necessity of assistance, COPD, ASHD, PVD, ischemic heart disease, and congestive heart disease, along with patient race, gender, and cubic functions of age and dialysis tenure. Facility controls include age, HRR, whether the facility is freestanding or hospital-based, and chain ownership status. BMI bin, GFR bin, HGB bin and income bin represent respectively the BMI quintile, GFR quintile, HGB quintile, and the income quintile of the ZIP code of residence for each patient. Additional controls include year-month fixed effects, facility fixed effects, and measures of the patient's health status relative to each measure ever used in the QIP during the patient's first six months in the data. Finally, where indicated, controls also include the patient's contemporary health status relative to each measure ever used in the QIP. More specifically, these controls take the form of indicators for each specific measure which take value 1 if the patient has readings which do not satisfy the QIP rules in the current period. Standard errors clustered by patient are in parentheses. *, ** and *** indicate significance at the 5%, 1% and 0.1%, respectively. Contemporary "Bad Vasc. Access," "Low Kt/V," and "High Calcium" are imputed to equal 0 prior to July 2010, January 2012 and May 2012, respectively. The regressions include indicator variables for the periods in which each of these measures is missing.

Table 4: Effort by QIP Measure

QIP Measure	Low HGB		High HGB		High Calcium	Low Kt/V	Transfusion	
Proxy for Effort	EPO	Iron	EPO	Iron	Cinacalcet	Time	EPO	Iron
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbb{1}\{\text{Penalty-Inducing}\}$	28.06***	17.89***	-9.062***	-11.00***	-0.0112***	-0.390	19.75***	11.94***
	(0.499)	(1.091)	(0.295)	(0.719)	(0.00332)	(0.285)	(0.501)	(1.324)
$\mathbb{1}\{\text{Penalty-Inducing}\} \times \mathbb{1}\{\text{Post}\}$	1.909**	4.642**	-2.038***	-2.571*	0.0578***	-1.221***	-13.52***	10.60***
	(0.677)	(1.540)	(0.402)	(1.023)	(0.00373)	(0.302)	(0.736)	(2.363)
Patient Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Facility FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month-Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Base Measures	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contemporary Measures	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.207	0.0967	0.207	0.0967	0.118	0.414	0.243	0.0559
Observations	2,422,033	2,422,033	2,422,033	2,422,033	8,587,554	8,907,135	12,836,157	12,836,157
Mean Dep. Var.	61.90	221.9	61.90	221.9	0.215	219.0	29.85	175.0

Notes: OLS estimates where the dependent variable for each QIP measure is a proxy for a facility's effort to improve a given QIP measure: EPOGEN doses, iron doses, an indicator variable that takes value 1 if cinacalcet is administered, and treatment time. For each QIP measure, $\mathbb{1}\{\text{Penalty-Inducing}\}$ indicates whether the patient failed to satisfy the QIP measure listed as the column title during October–December of the previous year. For example, in the columns labelled "Low HGB," $\mathbb{1}\{\text{Penalty-Inducing}\}$ is equal to 1 in year 2010 if a given patient has at least one low hemoglobin reading in the period October–December 2009, and 0 otherwise. $\mathbb{1}\{\text{Post}\}$ is an indicator variable that takes value 1 if the QIP measure is active in a given period. An observation is a patient-month. Sample consists of observations for in-center hemodialysis patients between the ages of 18 and 100 with Medicare as their primary payer from varying time periods: 2008–2010 for low HGB and high HGB to avoid the confounding effect of the switch from fee-for-service to prospective payments in 2011; 2012–2017 for high calcium and low Kt/V due to missing values from previous periods in the CROWNWeb data, which contain information on treatment time and calcium levels; and 2011–2017 for transfusion to avoid the confounding effect of the switch from fee-for-service to prospective payments in 2011. The first six months of observations for each patients are removed from the sample. Patient controls include indicator variables for incident comorbidities and characteristics reported on medical evidence forms, including diabetes, hypertension, BMI bin, GFR bin, HGB bin, income bin, high albumin, cancer, drug use, alcoholism, smoking behavior, necessity of assistance, COPD, ASHD, PVD, ischemic heart disease, and congestive heart disease, along with patient race, gender, and cubic functions of age and dialysis tenure. Facility controls include age, HRR, whether the facility is freestanding or hospital-based, and chain ownership status. BMI bin, GFR bin, HGB bin and income bin represent respectively the BMI quintile, GFR quintile, HGB quintile, and the income quintile of the ZIP code of residence for each patient. Additional controls include year-month fixed effects, facility fixed effects, and measures of the patient's health status relative to each measure ever used in the QIP during the patient's first six months in the data. Finally, where indicated, controls also include the patient's contemporary health status relative to each measure ever used in the QIP. More specifically, these controls take the form of indicator variables for each specific measure, which take value 1 if the patient has readings that do not satisfy the QIP rules in the current period. Standard errors clustered by patient are in parentheses. *, ** and *** indicate significance at the 5%, 1% and 0.1%, respectively. Contemporary "Bad Vasc. Access," "Low Kt/V," and "High Calcium" are imputed to equal 0 prior to July 2010, January 2012 and May 2012, respectively. The regressions include indicator variables for the periods in which each of these measures is missing.

Online Appendix

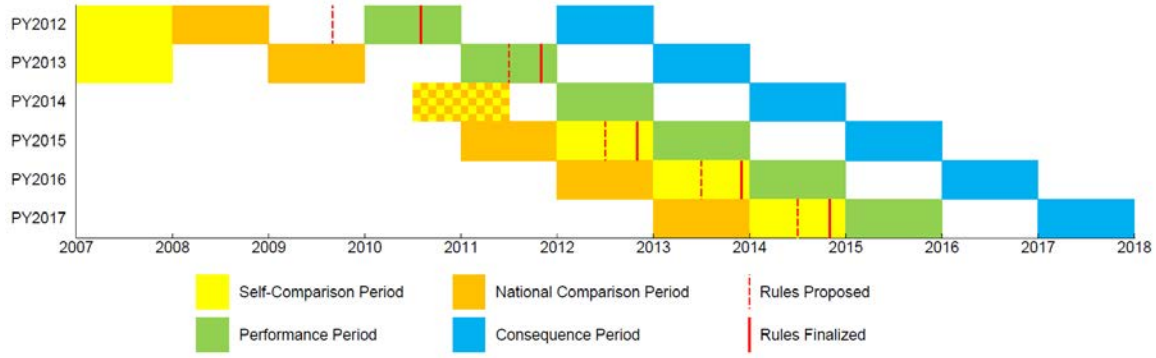
A Computing TPS Scores and QIP Timeline

Figure A1: Computing TPS Scores

PROJECTED PAYMENT REDUCTION PERCENTAGE		NO REDUCTION
TOTAL PERFORMANCE SCORE		
Facility Total Performance Score: 27 (out of 30)		
PERFORMANCE MEASURE SCORES		
Anemia Management Percent of Patients with hemoglobin less than 10 grams per deciliter (g/dL)		8 (out of 10)
Anemia Management Percent of patients with hemoglobin greater than 12 g/dL		10 (out of 10)
Dialysis Adequacy Percent of patients with urea reduction ratio (URR) of at least 65%		10 (out of 10)
Facility Rate Calculation for 2007		
7a	Number of patients with URR of at least 65%	49
7b	Total number of patients included in calculation	50
7c	Percent of patients with URR of at least 65% (Divide 7a by 7b and round)	98%
Performance Standard Determination		
7d	Facility comparison rate for 2007 (from 7c)	98%
7e	National average in 2008	96%
7f	Performance standard applied (Apply the lesser of 7d or 7e)	96%
Facility Rate Calculation for Performance Period (2010)		
7g	Number of patients with URR of at least 65%	34
7h	Total number of patients included in calculation	35
7i	Percent of patients with URR of at least 65% (Divide 7g by 7h and round)	97%
Performance Measure Score Calculation		
7j	Facility performance rate in 2010 (from 7i)	97%
7k	Performance standard (from 7f)	96%
7l	Does the facility meet or exceed the standard	Yes
7m	Difference between facility rate and performance standard-	Meets or Exceeds
7n	**Performance Measure Score**	**10**

Notes: The figure shows how the TPS score is computed from the QIP measures and claims data for a hypothetical facility. The Total Performance Score is based on a weighted average of the measure scores.

Figure A2: QIP Timeline



Notes: The figure presents the relevant events for each payment year from PY2012 to PY2017. The checkered yellow and orange period is both the self-comparison period and national comparison period.

B Summary Statistics for All Variables

Table A1: Summary Statistics — Patient Characteristics

	Observations	Mean	Std. Dev
Diabetic (%)	26,625,637	53.4	49.9
Hypertensive (%)	26,625,637	88.0	32.5
BMI (kg/m^2)	26,300,140	30.0	8.3
GFR	26,601,216	8.6	5.1
Dialysis Tenure (months)	26,625,907	53.2	46.2
HGB (g/dL)	26,625,907	9.7	1.9
High Albumin (%)	26,625,907	63.2	48.2
Cancer (%)	26,130,264	4.9	2.2
Drug use (%)	26,129,682	1.3	1.1
Drinker (%)	26,129,525	1.4	1.2
Smoker (%)	26,130,796	6.3	2.4
Needs Assistance (%)	26,134,413	9.3	2.9
COPD (%)	26,131,127	6.5	2.5
Atherosclerotic Heart Disease (%)	25,669,050	14.1	34.8
Peripheral Vascular Disease (%)	26,132,394	10.7	30.8
Ischemic Heart Disease (%)	26,130,706	8.1	27.3
Congestive Heart Failure (%)	26,136,434	27.5	44.6
Male (%)	26,625,090	54.5	49.8
Non-Hispanic White (%)	26,566,969	39.3	48.8
Black (%)	26,625,907	38.3	48.6
Hispanic (%)	26,566,969	16.6	37.2
Asian (%)	26,625,907	3.6	18.6
Other Race (%)	26,625,907	3.7	19.0

Notes: The table presents summary statistics for the incident clinical characteristics and patient demographics included in the regressions above. An observation is a patient-month. All variables are at incidence aside from dialysis tenure, which measures the number of months since the patient started dialysis . BMI, GFR, and HGB are included in the regressions in the form of indicator variables for each bin but reported here as continuous variables.

Table A2: Summary Statistics — Facility Characteristics

	Observations	Mean	Std. Dev
Median ZIP Income	58,658	49,445.3	19,555.2
Facility Age (years)	58,197	14.2	10.1
Chain (%)	58,197	80.4	39.7
Freestanding (%)	58,197	92.5	26.3
Facilities in ZIP	58,497	1.7	1.1
Mortality Rate	57,297	8.6	5.6
Hospitalization Rate	57,297	53.9	13.2
Infection Rate	57,297	12.2	6.9

Notes: The table presents summary statistics for facility characteristics and quality indicators. An observation is a facility-year. Median income by ZIP Code is included in the regressions in the form of indicator variables for each bin but reported here as a continuous variable. Mortality, hospitalization, and infection rates are measured as the percentage of patients who had at least one event during the year while at the facility. The denominator includes all patients who received dialysis at a facility for at least one month, so it includes switching patients (incoming or outcoming) as well.

Table A3: Summary Statistics — QIP Measures

	Observations	Mean	Std. Dev
Penalty-Inducing (%)	25,415,757	33.1	47.0
Penalty Score	24,813,002	0.09	0.15
Low HGB (%)	26,028,936	13.5	34.2
Low HGB (%)	26,028,936	13.5	34.2
High HGB (%)	26,028,936	11.0	31.3
Low URR (%)	25,815,787	4.9	21.6
Bad Vasc. Access (%)	20,268,446	37.1	48.3
Low Kt/V (%)	15,860,912	2.9	16.8
Infection (%)	26,625,907	1.9	13.6
Transfusion (%)	26,625,907	0.7	8.2
Readmission (%)	26,625,907	0.4	6.1
High Calcium (%)	14,249,025	2.6	16.0
Base Low HGB (%)	25,878,185	15.2	21.3
Base High HGB (%)	25,878,185	30.5	28.6
Base Low URR (%)	26,197,810	12.7	23.2
Base Bad Vasc. Access (%)	24,664,649	47.0	46.5
Base Low Kt/V (%)	22,607,867	4.3	12.9
Base Infection (%)	26,238,423	1.7	5.9
Base Transfusion (%)	26,238,423	0.2	2.1
Base Readmission (%)	26,238,423	0.3	2.3
Base High Calcium (%)	21,717,856	3.8	12.0

Notes: The table presents summary statistics for contemporary and baseline QIP measures. An observation is a patient-month. Low HGB indicates HGB below 10 g/dL. High HGB indicates HGB above 12 g/dL. Low URR indicates URR below 65%. Bad vascular access indicates receiving dialysis through a catheter or graft. Low Kt/V indicates Kt/V below 1.2. Readmission, infection, and transfusion indicate whether the patient was readmitted to the hospital within 90 days of a discharge, hospitalized for septicemia, or transfused in the month. Baseline versions of each measure are the patient-specific average of the measure over their first six months in the data.

Table A4: Summary Statistics — Measures of Facility Effort

	Observations	Mean	Std. Dev
EPOGEN (1000 IUs)	26,625,907	40.6	56.9
IV Iron (IUs)	26,625,907	202.7	245.4
Cinacalcet (%)	18,469,107	18.2	38.6
Treatment Time (minutes)	12,256,129	219.7	27.9

Notes: The table presents summary statistics for proxies of effort assessed in Table 4. An observation is a patient-month. Cinacalcet is an indicator variable for receiving any cinacalcet in the month. Treatment time is the reported time dialyzed in a patient's last dialysis session of the month.