Annals of Internal Medicine ORIGINAL RESEARCH

Changes in Dialysis Center Quality Associated With the End-Stage Renal Disease Quality Incentive Program

An Observational Study With a Regression Discontinuity Design

Kyle H. Sheetz, MD, MSc; Laura Gerhardinger, MA; Andrew M. Ryan, PhD; and Seth A. Waits, MD

Background: In 2012, the Centers for Medicare & Medicaid Services started levying performance-based financial penalties against outpatient dialysis centers under the mandatory End-Stage Renal Disease Quality Incentive Program.

Objective: To determine whether penalization was associated with improvement in dialysis center quality.

Design: Leveraging the threshold for penalization (total performance score < 60), a regression discontinuity design was used to examine the effect of penalization on quality over 2 years. Publicly available Medicare data from 2015–2018 were used. The effect of penalization at dialysis centers with different characteristics (for example, size or chain affiliation) was also examined.

Setting: United States.

Participants: Outpatient dialysis centers (n = 5830).

Measurements: Dialysis center total performance scores (a composite metric ranging from 0 to 100 based on clinical quality and adherence to reporting requirements) and individual measures that contribute to the total performance score.

Results: There were 1109 (19.0%) outpatient dialysis centers that received penalties in 2017 on the basis of performance in 2015. Penalized centers were located in ZIP codes with a higher average proportion of non-White residents (36.4% vs. 31.2%; P < 0.001) and residents with lower median income ($49 290 vs. $51 686; P < 0.001). Penalization was not associated with improvement in total performance scores in 2017 (0.4 point [95% CI, −2.5 to 3.2 points]) or 2018 (0.3 point [CI, −2.8 to 3.4 points]). This was consistent across dialysis centers with different characteristics. There was also no association between penalization and improvement in specific measures.

Limitation: The study could not account for how centers respond to penalization.

Conclusion: Penalization under the End-Stage Renal Disease Quality Incentive Program was not associated with improvement in the quality of outpatient dialysis centers.

Primary Funding Source: None.

Ann Intern Med. doi:10.7326/M20-6662 **Annals.org**
For author, article, and disclosure information, see end of text.
This article was published at Annals.org on 1 June 2021.

The quality of care at outpatient dialysis centers varies widely in the United States. Patients experience clinically meaningful differences in dialysis adequacy (as measured by the urea reduction ratio), a 2-fold difference in care satisfaction, and a nearly 5-fold difference in the likelihood of referral for kidney transplantation (1–4). In 2012, the Centers for Medicare & Medicaid Services (CMS) launched the End-Stage Renal Disease Quality Incentive Program (ESRD QIP), a novel mandatory pay-for-performance initiative designed to improve the quality of care at outpatient dialysis centers (5, 6). Under the program, centers face up to a 2% reduction in annual Medicare reimbursement if their performance on a range of quality measures falls below a certain threshold.

Despite its scope, the ESRD QIP has not been independently evaluated, and it remains unclear whether the quality of care at outpatient dialysis centers has improved as a result. Recent work suggests that patients treated at centers with lower program quality scores have a higher risk for death in their first year on dialysis (7). However, there is also evidence that dialysis centers are more likely to receive penalties if they are located in areas with lower household incomes or a higher proportion of ethnic minority residents, or in which more beneficiaries are dually enrolled in Medicare and Medicaid (8). This raises questions about whether the program is accurately measuring center quality or whether differences across centers are driven by underlying patient characteristics.

We used publicly available data from Medicare to evaluate whether penalization, one of several accountability mechanisms in the program, was associated with improvement in dialysis center quality from 2015 through 2018. In addition to evaluating overall changes in the program's total performance score, we examined whether the effect of penalization on quality varied at dialysis centers with different underlying characteristics or whether penalization was associated with improvement in specific quality measures.

METHODS

Study Design

We designed this analysis around penalties levied in calendar year 2017 on the basis of dialysis center performance in calendar year 2015. Under the ESRD QIP,

See also:

Web-Only
Supplement

centers received notification of these penalties in mid-2016. Notification occurs over a 1-month preview period during which centers receive details about their performance and can ask for clarification on how their scores were calculated. This analysis examined whether penalization was associated with changes in dialysis center performance in 2017, the year in which they were penalized, or in 2018, which allowed for additional time to respond to the penalty. This study was deemed exempt from review by the University of Michigan's institutional review board.

Data Sources

The CMS maintains a public registry with center-level data on specific quality measures, including those for the ESRD QIP, as part of Dialysis Facility Compare (9). For this analysis, we extracted data for calendar years 2015–2018. We also used these files to obtain information on dialysis center characteristics, including chain affiliation status, nonprofit status, and the number of dialysis stations. We excluded centers with incomplete data or those that did not record ESRD QIP total performance scores for any of the study years. Detailed information on cohort selection is shown in **Appendix Figure 1** (available at Annals.org). We also linked dialysis center quality data to ZIP code–level census data from the American Community Survey (2015–2018) on median household income and the proportion of non-White residents (10).

Outcomes

The primary outcome was the dialysis centers' total performance scores. This metric is based on performance across a range of clinical quality measures and adherence to program reporting requirements (11). The correlation between scores across study years and a detailed breakdown of each component of the total performance score, including annual averages for each measure, is provided in **Appendix Table 1** (available at Annals.org) and **Appendix Figure 2** (available at Annals.org). We also evaluated whether penalization was associated with changes in specific measures that were common across all study years. These included the proportion of patients receiving dialysis via arteriovenous fistula and catheters, the proportion of patients with hypercalcemia (serum calcium level >2.55 mmol/L [>10.2 mg/dL]), dialysis adequacy, the National Healthcare Safety Network (NHSN) standardized infection ratios for bloodstream infections, standardized readmission ratios, mineral metabolism, the In-Center Hemodialysis Consumer Assessment of Healthcare Providers and Systems Survey, and anemia management reporting (12). These outcomes were chosen because they were included in the total performance score across all study years and because they either reflect different, yet important, aspects of care provided at dialysis centers or are captured by different reporting mechanisms (for example, NHSN measures).

Statistical Analysis

We used the *t* test, Wilcoxon rank-sum test, or χ^2 test to compare characteristics of penalized and nonpenalized centers. We used a regression discontinuity design to evaluate the association between penalization and changes in dialysis center quality. Because we would not expect dialysis centers just above and below the penalization threshold (total performance score < 60) to differ in ways that would affect study outcomes, this approach mimics randomization, allowing us to isolate the effect of penalization on quality (13, 14). This quasi-experimental study design has other advantages. It is not subject to confounding from regression to the mean, is not contingent on counterfactual assumptions, and does not require an unexposed control population (15). This latter point is important because the ESRD QIP program is mandatory for all outpatient dialysis centers in the United States, thus limiting options for control selection. The **Appendix** (available at Annals.org) provides more detail on the analytical approach, including annotated code.

Using local linear regression, we fit separate models to estimate the effect of penalization on each study outcome (16). All estimates were derived by using the program rdrobust in Stata. Estimates were bias-corrected by using data-driven bandwidth selection and were reported with robust CIs (17). The CIs were derived by using bias correction, robust variance, and fixed-matches residuals. This approach has been previously shown to provide reliable estimates in regression discontinuity analyses (13, 16). The bandwidth specifies the range of total performance scores (and corresponding centers) included in each analysis. This is because regression discontinuity designs rely on the assumption that centers are effectively randomized just on either side of the penalty threshold. Thus, our estimates are local in that they are restricted to the data-driven bandwidths identified by the rdrobust program. We performed several stratified analyses to account for the fact that the effect of penalization may be modified by dialysis center characteristics. These included chain affiliations, nonprofit versus for-profit status, top versus bottom quartile size of the center (number of dialysis stations), top versus bottom quartile household income, and high versus low proportion of non-White minority residents. **Appendix Table 2** (available at Annals.org) shows specific data on the number of centers and the range of total performance scores used in each analysis.

We performed several sensitivity analyses. First, we replicated the main regression analysis while adjusting for dialysis center characteristics, including chain affiliation, for-profit status, number of dialysis stations, and patient demographic characteristics. It is also possible that centers would respond differently to, or be fundamentally different because of, repeated penalization in the years before 2015. To address this, we specified an alternative regression where we included only centers that received their first penalty in 2017 ($n = 5391$). Finally, to confirm the validity of our analytic approach, we estimated discontinuities in key dependent variables (number of dialysis stations, proportion of non-White minority residents). Should these falsification tests return a null finding, it would suggest that the centers are randomly distributed just above or below the penalization threshold.

We generated robust SEs accounting for clustering at the dialysis center level and used a 2-sided threshold for significance of less than 0.05. All analyses were performed in Stata, version 15.

Role of the Funding Source

There was no funding associated with this study.

RESULTS

Characteristics of Outpatient Dialysis Centers

Of 5830 outpatient dialysis centers, 1109 (19.0%) received penalties in 2017 (Appendix Figures 2 and 3, available at Annals.org). Among penalized centers, 931 were chain-affiliated (84.0%), compared with 4418 nonpenalized centers (93.6%) ($P < 0.01$). A similar proportion of penalized and nonpenalized centers (986 [88.9%] vs. 4283 [90.7%], respectively; $P = 0.07$) had a for-profit business model (Table 1). Penalized centers (36.4%) were located in ZIP codes with a higher average proportion of non-White minority patients compared with nonpenalized centers (31.2%) ($P < 0.01$). The median annual income at the ZIP code level was also lower for penalized centers compared with nonpenalized centers ($49 290 [interquartile range, $36 339 to $58 284]) vs. $51 686 [interquartile range, $37 974 to $61 409], respectively; $P < 0.01$).

Relationship Between Penalization and Total Performance Scores

Penalization in 2017 was not associated with improvement in dialysis centers' total performance scores in 2017 (0.4 point [95% CI, −2.5 to 3.2 points]) or 2018 (0.3 point [CI, −2.8 to 3.4 points]) (Table 2 and the Figure). These estimates were similar after adjustment for dialysis center characteristics and when analysis was restricted to centers newly penalized in 2017 (Table 2 and Supplement Figures 1 and 2, available at Annals.org). Penalization was also not associated with improvement in total performance scores in analyses stratified by different types of dialysis centers (Table 2 and Supplement Figures 3 to 12, available at Annals.org). For example, penalization was not associated with improvement in total performance scores in either year regardless of business model or size of the dialysis center: 0.6 point (CI, −2.3 to 3.5 points) in 2017 and 0.4 point (CI, −2.7 to 3.5 points) in 2018 at for-profit centers; 0.7 point (CI, −7.2 to 8.6 points) and −2.3 points (CI, −13.7 to 9.2 points), respectively, at nonprofit centers; −0.7 point (CI, −6.5 to 5.2 points) and 1.5 points (CI, −3.7 to 6.6 points) at large centers; and −1.3 points (CI, −7.9 to 5.2 points) and −1.5 point (CI, −8.5 to 5.6 points) at large centers.

Relationship Between Penalization and Specific Quality Measures

There was no significant association between penalization vand improvement in specific components of the total

Table 1. Characteristics of Outpatient Dialysis Centers in the United States*

Characteristic	All Centers (*n* = 5830)	Nonpenalized Centers (*n* = 4721)	Penalized Centers (*n* = 1109)
Dialysis center			
Dialysis stations, *n*			
Mean (SD)	18 (8)	18 (8)	19 (9)
Median (IQR)	18 (12–24)	17 (12–23)	19 (13–24)
Chain affiliation, *n (%)*	5349 (91.8)	4418 (93.6)	931 (84.0)
For-profit, *n (%)*	5269 (90.4)	4283 (90.7)	986 (88.9)
Local population demographic			
Non-White minority (SD), %†	32.2 (24.0)	31.2 (23.6)	36.4 (25.5)
Median annual income (IQR), $‡	51231 (37 732–60 781)	51686 (37 974–61 409)	49290 (36 339–58 284)
U.S. Census region			
Northeast	767 (13.2)	607 (12.9)	160 (14.4)
Midwest	1317 (22.6)	1058 (22.4)	259 (23.4)
South	2680 (46.0)	2101 (44.5)	579 (52.2)
West	1066 (18.3)	955 (20.2)	111 (10.0)
Mean ESRD QIP outcomes in 2015 (SD)			
Total performance score	68.8 (11.1)	72.6 (8.1)	52.7 (6.9)
Arteriovenous fistula utilization rate§	66.2 (10.7)	67.8 (10.2)	59.6 (10.4)
Catheter utilization rate‖	10.0 (6.5)	8.9 (5.7)	14.3 (7.8)
Kt/V dialysis adequacy¶	7.3 (2.3)	7.8 (1.8)	5.3 (2.9)
NHSN bloodstream infections**	1.0 (0.8)	0.9 (10.7)	1.4 (0.9)
Standardized readmission ratio††	1.0 (0.3)	0.9 (0.3)	1.2 (0.2)
Hypercalcemia‡‡	1.3 (1.4)	1.1 (1.2)	2.0 (1.7)
ICH CAHPS¶	9.8 (1.4)	9.9 (0.8)	9.1 (2.8)
Anemia management¶	9.8 (0.6)	9.9 (0.5)	9.8 (1.0)

ESRD QIP = End-Stage Renal Disease Quality Incentive Program; ICH CAHPS = In-Center Hemodialysis Consumer Assessment of Healthcare Providers and Systems Survey; IQR = interquartile range; NHSN = National Healthcare Safety Network.
* Data are derived from dialysis center data contained in the Dialysis Facility Compare files maintained by the Centers for Medicare & Medicaid Services.
† The proportion of non-White residents within the same ZIP code based on 2017 census data.
‡ Annual income represents the median income of citizens in 2017 in the same zip code as the dialysis center. All other data on dialysis centers is from 2017, the first year of outcomes assessed in the study.
§ Percentage of patient-months on dialysis during the last hemodialysis session of the month accessing an autogenous fistula with 2 needles.
‖ Percentage of patient-months on dialysis during the last hemodialysis session of the month accessing a catheter continuously for 90 days or longer.
¶ A program-specific score ranging from 0–10.
** Ratio of the number of new positive blood cultures drawn as an outpatient to the number of maintenance in-center hemodialysis patients treated by the center.
†† Ratio of the number of observed unplanned 30-day hospital readmissions to the number of expected unplanned 30-day readmissions.
‡‡ Percentage of patient-months with a 3-month rolling average of total uncorrected serum calcium levels >2.55 mmol/L (>10.2 mg/dL).

Table 2. Changes in Total Performance Scores Associated With Penalization Under the ESRD QIP

Variable	Total Program Performance Scores*				
Type of dialysis center	Baseline Score in 2015	2017		2018	
		Change in Score (95% CI)	Centers, *n*	Change in Score (95% CI)	Centers, *n*
All centers	68.8 (11.1)	0.4 (−2.5 to 3.2)	1907	0.3 (−2.8 to 3.4)	1907
All centers, risk-adjusted†	68.8 (11.1)	0.8 (−2.0 to 3.5)	1905	0.4 (−2.5 to 3.3)	2179
Newly penalized centers‡	69.5 (10.6)	0.5 (−2.3 to 3.3)	1745	0.2 (−3.1 to 3.5)	1745
For-profit centers	68.9 (11.1)	0.6 (−2.3 to 3.5)	1740	0.4 (−2.7 to 3.5)	1986
Nonprofit centers	68.5 (11.2)	0.7 (−7.2 to 8.6)	195	−2.3 (−13.7 to 9.2)	167
Chain-affiliated centers	69.4 (10.6)	0.4 (−2.5 to 3.4)	1747	−0.3 (−3.2 to 2.6)	2219
Non–chain-affiliated centers	62.9 (15.0)	0.9 (−6.5 to 8.3)	207	4.4 (−5.5 to 14.3)	183
Large centers§	67.0 (9.8)	−0.7 (−6.5 to 5.2)	444	1.5 (−3.7 to 6.6)	635
Small centers§	70.8 (13.4)	−1.3 (−7.9 to 5.2)	466	−1.5 (−8.5 to 5.6)	466
Centers in minority-predominant ZIP codes‖	66.9 (10.9)	2.2 (−4.3 to 8.7)	460	2.8 (−4.5 to 10.1)	567
Centers in not minority-predominant ZIP codes	70.3 (11.6)	4.6 (−2.2 to 11.4)	427	1.6 (−5.9 to 9.1)	427
Centers in high-income ZIP codes¶	69.3 (10.9)	0.1 (−4.9 to 5.1)	451	3.0 (−3.1 to 9.0)	451
Centers in low-income ZIP codes	67.9 (10.7)	−0.3 (−6.6 to 6.1)	440	−1.8 (−8.5 to 4.8)	440
Measure	**Performance on Specific Measures***				
	Baseline Score in 2015	2017		2018	
		Change in Score (95% CI)	Centers, *n*	Change in Score (95% CI)	Centers, *n*
Arteriovenous fistula utilization rate	66.2 (10.7)	1.4 (−1.1 to 3.8)	2114	1.0 (−1.4 to 3.4)	2354
Catheter utilization rate	10.0 (6.5)	−0.9 (−1.8 to 1.6)	1853	−0.1 (−1.6 to 1.4)	2121
Kt/V dialysis adequacy	1.3 (1.4)	0.1 (−0.2 to 0.5)	1554	−0.1 (−0.4 to 0.3)	1553
NHSN bloodstream infections	7.3 (2.3)	0.2 (−0.3 to 0.8)	1904	0.1 (−0.5 to 0.7)	2177
Standardized readmission ratio	1.0 (0.8)	−0.1 (−0.2 to 0.2)	1803	−0.1 (−0.2 to 0.1)	1819
Hypercalcemia	9.8 (1.4)	−0.6 (−1.4 to 0.2)	1100	−0.6 (−1.6 to 0.4)	1003
ICH CAHPS	1.0 (0.3)	−0.002 (−0.1 to 0.1)	1903	0.02 (−0.04 to 0.1)	1903
Anemia management	9.8 (0.6)	0.1 (−0.2 to 0.3)	1553	0.1 (−0.1 to 0.3)	1551

ESRD QIP = End-Stage Renal Disease Quality Incentive Program; ICH CAHPS = In-Center Hemodialysis Consumer Assessment of Healthcare Providers and Systems Survey; NHSN = National Healthcare Safety Network.

* The effect of penalization can be interpreted as the independent per-point effect of program penalization on centers' total performance scores ranging from 0 to 100. Estimates are derived from regression discontinuity models with data-driven bandwidth selection and robust CIs. The number of centers reflects the total number of centers within the regression discontinuity analysis bandwidth. More specific data on the number of centers and scores that define each bandwidth are included in Appendix Table 2.

† Models were adjusted for dialysis center characteristics, including for-profit status, chain affiliation, and size (measured as the number of dialysis stations). Models were also adjusted for local population demographic characteristics measured at the ZIP code level, including proportion of non-White minority residents and median household income.

‡ Newly penalized centers (n = 5391) were those that did not receive a penalty under the ESRD QIP program in the 2 years before 2017.

§ Large centers represent the top quartile (>24 stations); small centers represent the bottom quartile (<12 stations).

‖ Centers were stratified by the proportion of non-white residents within the same zip code based on 2017 census data. Minority predominant zip codes represent the top quartile (range 46.4% to 99.6%), where not minority-predominant represent the bottom quartile (range, 0.9% to 12.7%).

¶ Centers were stratified by the median household income within the same zip code based on 2017 census data. High income zip codes represent the top quartile (range $60 781 to $183 656), where low income represent the bottom quartile (range, $9409 to $37 732).

All baseline performance reflects the mean and standard deviation for the centers meeting the inclusion criteria for each analysis. Overall baseline statistics for all centers are included in Table 1.

performance score (Table 2 and Supplement Figures 13 to 20, available at Annals.org). For example, penalization was not associated with improvement in arteriovenous fistula utilization rates (1.4 percentage points [CI, −1.1 to 3.8 percentage points] in 2017 and 1.0 percentage point [CI, −1.4 to 3.4 percentage points] in 2018) or anemia management reporting (0.1 percentage point [CI, −0.2 to 0.3 percentage point] in 2017 and 0.1 percentage point [CI, −0.1 to 0.3 percentage point] in 2018).

Falsification Tests

In our falsification tests, there was no association between penalization and changes in the number of dialysis stations (1.0 [CI, −0.9 to 3.0]) or proportion of

non-White minority residents (−0.2 percentage point [CI, −6.0 to 5.5 percentage points]) (**Supplement Table**, available at Annals.org).

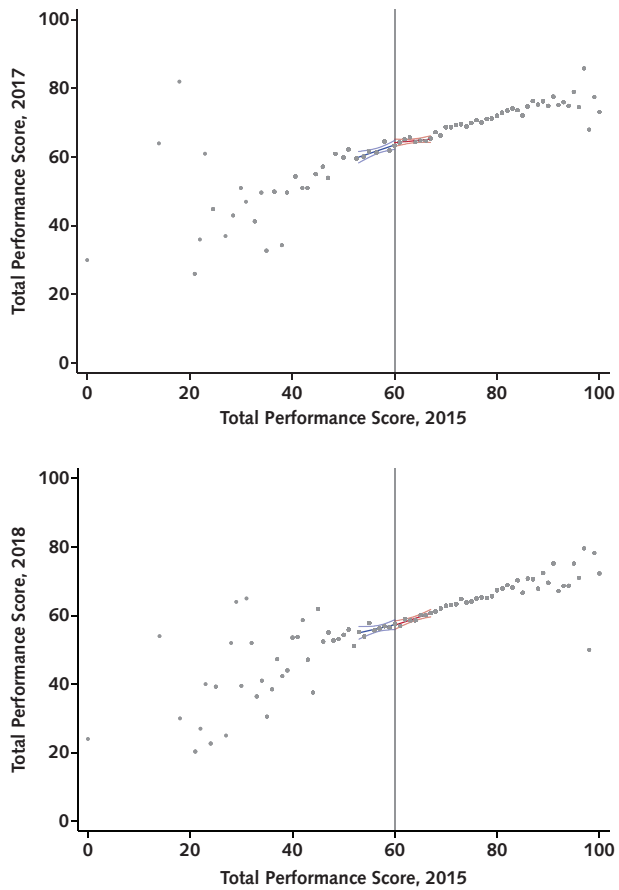
Discussion

This evaluation of the ESRD QIP suggests that penalization was not associated with improvement in dialysis center quality. This was consistent across dialysis centers with different business models, chain affiliation status, and number of dialysis stations, and for centers that serve underserved populations and are more likely to receive penalties under the program. There was also no association between penalization and centers' improvement in specific clinical quality or reporting measures included in the program's total performance score.

Outpatient or long-term maintenance dialysis is an important clinical resource and national health care policy priority. The CMS spent $11.4 billion in 2017 on outpatient dialysis services for approximately 500 000 beneficiaries with ESRD on dialysis (18). But spending and the overall quality of care vary widely across dialysis centers or different regions of the country (19, 20). Previous research suggests that attention to particular domains of ESRD care, such as increasing dialysis adequacy (measured by the urea reduction ratio) or reducing the incidence of anemia, is associated with lower mortality and higher quality of life (21, 22). Although these and other measures are included in the ESRD QIP, prior evidence on whether the program itself has improved outpatient dialysis care is limited. Most previous work has instead focused on characteristics of penalized centers. For example, cross-sectional analyses suggest that larger centers, those affiliated with a dialysis chain, and those located in the South or Northeastern parts of the United States achieve higher quality scores and, as a result, are less likely to receive penalties (23). However, similar analyses from other program years found the opposite: Larger dialysis centers, for instance, had lower quality scores (24).

Our study goes beyond prior work to address the fundamental question of whether dialysis center quality in the United States has improved under the ESRD QIP. Using the threshold for penalties to isolate program effects, our study overcomes a critical limitation in evaluating the ESRD QIP: Participation is mandatory, which makes it impossible to otherwise compare outcomes against control centers not exposed to program penalties. Moreover, our findings highlight several potential issues with the design of the ESRD QIP. First, the program uses a broad list of measures that reflect a range of quality priorities. Thus, they may not permit centers to focus sufficient efforts to address each measure independently. This aligns with concerns that annual penalties are based on a growing number of measures that change frequently (25). Second, because the measures do change frequently, it is possible for centers to receive penalties in a given year for their performance on measures no longer incentivized by the program. We found that a greater proportion of nonpenalized centers happened to be chain-affiliated. This observation may suggest that larger chains with greater infrastructure are better equipped to manage the frequent changes in quality benchmarks and data reporting. If this is true, penalization may be more closely associated with the structural capabilities of the center rather than clinical quality. Third, the program issues annual penalties. This further assumes that centers can implement effective changes on a yearly timescale. These latter points may also be generalizable to other pay-for-performance programs, where changes in quality metrics and complex penalization schemes could limit effectiveness. Finally, the ESRD QIP may levy a disproportionate share of penalties on certain types of centers. Because many of these centers care for poor and otherwise underserved communities, they may exacerbate existing health care disparities (7, 8, 24).

Figure. Association between End-Stage Renal Disease Quality Incentive Program penalization in 2015 and total performance scores in 2017 (*top*) and 2018 (*bottom*).



Scores range from 0 to 100. The vertical line identifies the threshold for penalization based on performance in 2015 (total performance score, 60). Dots represent the mean total performance scores in 2017 or 2018 for all centers that achieved a particular total performance score in 2015. These mean values display the full range of total performance scores for all centers, not only those included in the regression discontinuity analysis. The bandwidths for the regression discontinuity analysis for both 2017 and 2018 include centers with scores of 53 to 67. Trend lines and CIs reflect the only centers sufficiently close to the threshold to be included in the regression discontinuity model (that is, the model bandwidth). Trend lines were derived from local linear regression with associated robust CIs. The blue lines reflect the trend below the penalization threshold, and the red lines reflect the trend above. Discontinuity estimates for the effect on total performance scores were 0.4 point (95% CI, −2.5 to 3.2 points) in 2017 and 0.3 point (CI, −2.8 to 3.4 points) in 2018.

That said, there are opportunities to improve the program. Certain changes could be made under the federal rule-making process, which may increase the efficiency in which they are implemented. For example, CMS could commit to a series of measures in 3- or 5-year increments. Increasing the predictable time of inclusion may allow centers to plan and respond more effectively. Because the full complement of measures that contributed to performance scores differed somewhat between 2015, 2017, and 2018, it is possible that our main analysis may not be a fair comparison of centers' performance across the study period. To further address this, we examined whether centers improved in response to penalties for measures that were included in the total performance scores for each year. We found similar results, suggesting that penalized centers did not improve under both the global program performance measure and under specific measures of quality. The CMS could also reduce the number of measures and instead focus on those with the most variation or direct effects on patients' mortality, quality of life, and access to kidney transplantation. This may also provide clinicians with more proximate control over the quality of care delivered, rather than diffuse measures that providers have little ability to influence. Other changes to the program would require legislative action, such as changes to the annual structure of penalties to allow times during which centers are not subject to payment reductions. The CMS could also consider increasing the financial penalties. Even a 2% reduction in reimbursement may be insufficient to motivate dialysis centers to change practices. However, this would also increase the extent to which the quality measures are accurate because penalization may have a broad negative effect on patients treated at certain centers. Adjusting measures for social risk factors, such as dual-eligibility or median household income, may also help reduce the chances that program penalties increase existing disparities (26, 27).

Our study has limitations. First, it is possible that our overall analytic approach is invalid if centers above and below the penalization threshold are measurably different from one another. Our falsification tests that estimated discontinuities for key dependent variables, including dialysis center size and local proportion of non-White residents, do not suggest that this is the case. Although it is possible that our overall findings mask the fact that certain types of centers were able to effectively respond to the penalties, our sensitivity analyses both adjusted for and were stratified by these factors yet demonstrated no evidence of improvement associated with penalization. That said, it is possible that the simple existence of policy has an effect on dialysis center quality. Second, with mandatory participation, we are unable to account for the possibility that quality would be worse overall in its absence.

Third, a 2-year period may not be sufficient to capture improvement in quality related to the penalties. That quality could manifest as real changes in QIP metrics or from broader outcomes such as improvement in mortality. Even if this is true, our findings are still important because the ESRD QIP suggests by design that improvement can occur on an annual basis given the timeline for performance and penalization. Moreover, we do not know how specific centers responded to the program. These data will be critical to future changes in the design of the QIP.

Finally, patient factors may affect our findings. Although we did not adjust for patient-level factors, we accounted for local population demographic characteristics known to affect ESRD QIP total performance scores. Moreover, because we used the same data Medicare uses to leverage penalties under the program, we accounted for the same inclusion and exclusion criteria and patient-time at risk for denominators used by the program.

In conclusion, we found that performance-based financial penalties under the ESRD QIP were not associated with improvement in the quality of outpatient dialysis centers. This was consistent across a range of different types of centers and individual quality metrics included in the program's total performance score. These data suggest that CMS may consider changes to the program design as they continue to experiment with ways to improve the care of patients with ESRD.

From University of Michigan, Center for Healthcare Outcomes and Policy, and Center for Evaluating Health Reform, Ann Arbor, Michigan (K.H.S.); Center for Healthcare Outcomes and Policy, Ann Arbor, Michigan (L.G.); Center for Healthcare Outcomes and Policy, Center for Evaluating Health Reform, and University of Michigan School of Public Health, Ann Arbor, Michigan (A.M.R.); and University of Michigan and Center for Healthcare Outcomes and Policy, Ann Arbor, Michigan (S.A.W.).

Disclosures: Authors have reported no disclosures of interest. Forms can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M20-6662.

Reproducible Research Statement: *Study protocol:* Available from Dr. Sheetz (e-mail, ksheetz@med.umich.edu). *Statistical code:* See the **Appendix** (available at Annals.org). *Data set:* Available at www.medicare.gov/care-compare.

Corresponding Author: Kyle H. Sheetz, MD, MSc, Center for Healthcare Outcomes and Policy, 2800 Plymouth Road, NCRC B016, Room 100N-11, Ann Arbor, MI 48109; e-mail, ksheetz@med.umich.edu.

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

References

1. **Brady BM, Zhao B, Niu J, et al.** Patient-reported experiences of dialysis care within a national pay-for-performance system. JAMA Intern Med. 2018;178:1358-1367. [PMID: 30208398] doi:10.1001/jamainternmed.2018.3756

2. **Fink JC, Zhan M, Blahut SA, et al.** Measuring the efficacy of a quality improvement program in dialysis adequacy with changes in center effects. J Am Soc Nephrol. 2002;13:2338-44. [PMID: 12191978]

3. **Paul S, Plantinga LC, Pastan SO, et al.** Standardized transplantation referral ratio to assess performance of transplant referral among dialysis facilities. Clin J Am Soc Nephrol. 2018;13:282-289. [PMID: 29371341] doi:10.2215/CJN.04690417

4. **Saunders MR, Chin MH.** Variation in dialysis quality measures by facility, neighborhood, and region. Med Care. 2013;51:413-7. [PMID: 23579351] doi:10.1097/MLR.0b013e318287d720

5. **Weiner D, Watnick S.** The ESRD Quality Incentive Program—can we bridge the chasm. J Am Soc Nephrol. 2017;28:1697-1706. [PMID: 28298324] doi:10.1681/ASN.2016101079

6. **Centers for Medicare & Medicaid Services (CMS), HHS.** Medicare program; end-stage renal disease quality incentive program. Final rule. Fed Regist. 2011;76:627-46. [PMID: 21261127]
7. **Ajmal F, Probst JC, Brooks JM, et al.** Freestanding dialysis facility quality incentive program scores and mortality among incident dialysis patients in the United States. Am J Kidney Dis. 2020;75:177-186. [PMID: 31685294] doi:10.1053/j.ajkd.2019.07.023
8. **Qi AC, Butler AM, Joynt Maddox KE.** The role of social risk factors in dialysis facility ratings and penalties under a Medicare quality incentive program. Health Aff (Millwood). 2019;38:1101-1109. [PMID: 31260369] doi:10.1377/hlthaff.2018.05406
9. **Centers for Medicare & Medicaid Services.** Dialysis facility comparison tool. Accessed at www.medicare.gov/dialysisfacilitycompare/#qip/quality-incentive-program on 11 January 2020.
10. **U.S. Census Bureau.** American FactFinder—community facts. 2015-2018. Accessed at www.census.gov/programs-surveys/acs on 25 April 2019.
11. **Centers for Medicare & Medicaid Services.** ESRD Quality Incentive Program. 11 February 2020. Accessed at www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/ESRDQIP on 21 April 2020.
12. **Centers for Medicare & Medicaid Services.** End-Stage Renal Disease (ESRD) Quality Incentive Program (QIP) Payment Year (PY) 2017 Final Measure Technical Specifications. Centers for Medicare & Medicaid Services; 2016:1-17.
13. **Sankaran R, Sukul D, Nuliyalu U, et al.** Changes in hospital safety following penalties in the US Hospital Acquired Condition Reduction Program: retrospective cohort study. BMJ. 2019;366:l4109. [PMID: 31270062] doi:10.1136/bmj.l4109
14. **Desai S, McWilliams JM.** Consequences of the 340B Drug Pricing Program. N Engl J Med. 2018;378:539-548. [PMID: 29365282] doi:10.1056/NEJMsa1706475
15. **Venkataramani AS, Bor J, Jena AB.** Regression discontinuity designs in healthcare research. BMJ. 2016;352:i1216. [PMID: 26977086] doi:10.1136/bmj.i1216
16. **Calonico S, Cattaneo MD, Titunik R.** Robust data-drive inference in the regression-discontinuity design. Stata J. 2014;14:909-46.
17. **Calonico S, Cattaneo MD, Farrel MH, et al.** rdrobust: software for regression-discontinuity designs. Stata J. 2017;17:372-404.
18. **Medicare Payment Advisory Commission.** Outpatient dialysis services. In: Report to the Congress: Medicare Payment Policy. March 2019. Accessed at www.medpac.gov/docs/default-source/reports/mar19_medpac_entirereport_sec.pdf?sfvrsn=0 on 30 April 2021.
19. **Hirth RA, Tedeschi PJ, Wheeler JR.** Extent and sources of geographic variation in Medicare end-stage renal disease expenditures. Am J Kidney Dis. 2001;38:824-31. [PMID: 11576886]
20. **Rodriguez RA, Sen S, Mehta K, et al.** Geography matters: relationships among urban residential segregation, dialysis facilities, and patient outcomes. Ann Intern Med. 2007;146:493-501. [PMID: 17404351]
21. **McClellan WM, Soucie JM, Flanders WD.** Mortality in end-stage renal disease is associated with facility-to-facility differences in adequacy of hemodialysis. J Am Soc Nephrol. 1998;9:1940-7. [PMID: 9773796]
22. **Kliger AS, Finkelstein FO.** Can we improve the quality of life for dialysis patients? [Editorial]. Am J Kidney Dis. 2009;54:993-5. [PMID: 19932876] doi:10.1053/j.ajkd.2009.09.005
23. **Ajmal F, Probst J, Brooks J, et al.** Association between freestanding dialysis facility size and Medicare Quality Incentive Program performance scores. Am J Nephrol. 2019;49:64-73. [PMID: 30557871] doi:10.1159/000495262
24. **Saunders MR, Lee H, Chin MH.** Early winners and losers in dialysis center pay-for-performance. BMC Health Serv Res. 2017;17:816. [PMID: 29216894] doi:10.1186/s12913-017-2764-4
25. **Diamond LH, Howard AD.** The ESRD Quality Incentive Program: the current limitations of evidence and data to develop measures, drive improvement, and incentivize outcomes. Adv Chronic Kidney Dis. 2016;23:377-384. [PMID: 28115082] doi:10.1053/j.ackd.2016.11.007
26. **Johnston KJ, Joynt Maddox KE.** The role of social, cognitive, and functional risk factors in Medicare spending for dual and nondual enrollees. Health Aff (Millwood). 2019;38:569-576. [PMID: 30933581] doi:10.1377/hlthaff.2018.05032
27. **Joynt Maddox KE, Reidhead M, Hu J, et al.** Adjusting for social risk factors impacts performance and penalties in the Hospital Readmissions Reduction Program. Health Serv Res. 2019;54:327-336. [PMID: 30848491] doi:10.1111/1475-6773.13133

Current Author Addresses: Drs. Sheetz, Ryan, and Waits and Ms. Gerhardinger: Center for Healthcare Outcomes and Policy, 2800 Plymouth Road, NCRC B016, Room 100N-11, Ann Arbor, MI 48109.

Author Contributions: Conception and design: K.H. Sheetz, A.M. Ryan, S.A. Waits.
Analysis and interpretation of the data: K.H. Sheetz, L. Gerhardinger, A.M. Ryan.
Drafting of the article: K.H. Sheetz, A.M. Ryan, S.A. Waits.
Critical revision for important intellectual content: K.H. Sheetz, A.M. Ryan, S.A. Waits.
Final approval of the article: K.H. Sheetz, L. Gerhardinger, A.M. Ryan, S.A. Waits.
Statistical expertise: L. Gerhardinger, A.M. Ryan.
Administrative, technical, or logistic support: S.A. Waits.
Collection and assembly of data K.H. Sheetz, L. Gerhardinger.

Appendix: Methods and Technical Specifications

We performed a regression discontinuity analysis to address issues of selection bias that may limit our ability to determine whether program penalties themselves were independently associated with future performance under the ESRD QIP. The regression discontinuity design capitalizes on the idea that centers just above and below the penalty threshold should in theory be similar or "randomly" allocated about the threshold. Thus, differences in the study outcomes may be less likely to be biased by any factors (such as patient demographic characteristics) and are therefore attributable to the policy itself.

Data-Driven Bandwidth Selection

When performing regression discontinuity analysis, the researcher must choose the bandwidth, which defines the sample that will be included in the region around the threshold value. For instance, if the discontinuity cutoff was at 50 for the continuous variable used to identify the discontinuity (2015 total performance score in this study) and the bandwidth was 5, then the sample would include observations between 45 and 55 on that continuous variable.

The choice of bandwidth is important because observations far from the threshold may hold little information about the effects of changes near the threshold and because results can be sensitive to the inclusion of observations around the cutoff. Use of larger bandwidths tends to decrease statistical variance at the expense of requiring more assumptions for statistical inference (28).

To address this problem, researchers have developed tools to choose bandwidth on the basis of statistical criteria designed to maximize the robustness of estimation to violations in statistical assumptions. We used the bandwidth selection routine created by Catteno and colleagues (17), which minimizes the joint combination of estimator bias and variance.

Analytic Code for Regression Discontinuity Analyses

Main Effects Analysis

rdrobust totalperformancescore2017 totalperformancescore2015, c(60) all

Where 2017 and 2015 denote the year of total performance score for each center included in the analysis. The code "c(60)" reflects the discontinuity threshold for penalization in 2015.

The "all" command returns estimates from 3 different procedures: conventional regression discontinuity estimates with conventional variance estimator, bias-corrected regression discontinuity estimates with conventional variance estimator, and bias-corrected regression discontinuity estimates with robust variance estimator. We report from the bias-corrected regression discontinuity estimates with robust variance estimator for each model specification.

Risk-Adjusted Analysis

rdrobust totalperformancescore2017 totalperformancescore2015, covs(chain forprofit size income nonwhite male) c(60) all

Where "chain" and "forprofit" reflect categorical covariates for chain affiliation and for-profit status, respectively; and "size," "income," "nonwhite," and "male" reflect continuous variables for the number of dialysis stations, median income, proportion of non-white residents, and proportion male, respectively.

Stratified Analysis (Example)

rdrobust totalperformancescore2017 totalperformancescore2015 if chain == 1, c(60) all

Where we restrict to only those centers that are chain affiliated.

We used a combination of rdplot and twoway plot to generate all regression discontinuity figures.

Figure Code (Example)

rdplot totalperformancescore2017 totalperformancescore2015, c(60)

kernal(triangular) p(1) graph_options(title("") ytitle(""{bf:Total Performance Score, 2017}"", margin(small)) yscale (range(0 100)) ylabel(0(20)100) xtitle("{bf:Total Performance Score, 2015}", margin(small)) legend(off) graphregion (color(white))) genvars

Where 2017 and 2015 denote the year of total performance score for each center included in the analysis. The code "c(60)" reflects the discontinuity threshold for penalization in 2015.

sort totalperformancescore2015

keep rdplotoutput rdplot_hat_y totalperformancescore2015 totalperformancescore2017 rdplot_mean_x rdplot_mean_y

gen above_threshold = totalperformancescore2017 if totalperformancescore2015>= 60 & totalperformancescore2015<= 67 /*Edit here for each graph*/

gen below_threshold = totalperformancescore2017 if totalperformancescore2015<= 60 & totalperformancescore2015>= 53 /*Edit here for each graph*/

The thresholds above define the bandwidth for plotting the trend lines. Below is the remainder of the figure.

twoway (lfitci above_threshold totalperformancescore 2015, ciplot(rline) color(cranberry)) (lfitci below _threshold totalperformancescore2015, ciplot(rline)

color(blue)) (scatter rdplot_mean_y rdplot_mean_x, yscale(range(0100)) ylabel(0(20)100) mcolor(gs8) msize (vsmall)), xline(60, lcolor(gs8)) title("") ytitle(""{bf:Total Performance Score, 2017}"', margin(small)) xtitle("{bf: Total Performance Score, 2015}", margin(small)) legend (off) graphregion(color(white)) play(RDPlot_format)

Web Reference

28. **Cattaneo MD, Vazques-Bare G.** The choice of neighborhood in regression discontinuity designs. Obs Stud. 2016;2:134-46.

Appendix Table 1. Composition of Total Performance Scores, 2015–2018*

Performance Score Component	2015			2016			2017				2018			
	Mean Score (SD)	Clinical (75%)	Reporting (25%)	Mean Score (SD)	Clinical (80%)	Reporting (10%)	Mean Score (SD)	Clinical (75%)	Safety (15%)	Reporting (10%)	Mean Score (SD)	Clinical (75%)	Safety (15%)	Reporting (10%)
Total score	68.8 (11.1)	NA	NA	62.4 (12.6)	NA	NA	67.1 (12.2)	NA	NA	NA	61.7 (13.6)	NA	NA	NA
Arteriovenous fistula utilization rate	5.6 (2.6)	+		5.1 (2.6)	+		5.3 (2.6)	+			5.2 (2.7)	+		
Kt/V dialysis adequacy	7.3 (2.3)	+		7.4 (2.4)	+		8.2 (2.1)	+			7.3 (2.6)	+		
NHSN bloodstream infections	5.2 (3.0)	+		5.5 (3.0)	+		7.3 (2.0)		+		7.4 (1.9)		+	
Standardized readmission ratio	4.8 (3.2)	+		4.7 (2.9)	+		4.8 (3.0)	+			4.7 (3.0)	+		
Hypercalcemia	7.6 (2.3)	+		8.0 (2.0)	+		8.3 (2.1)	+			6.8 (2.8)	+		
ICH CAHPS	9.8 (1.4)		+	4.8 (2.6)	+		5.2 (2.5)	+			5.0 (2.6)	+		
Mineral metabolism	8.9 (1.7)		+	9.0 (1.7)		+	9.1 (1.6)			+	–			
Anemia management	9.8 (0.6)		+	9.9 (0.6)		+	9.9 (0.5)			+	9.9 (0.4)			+
Standardized transfusion ratio	–			5.3 (3.1)	+		5.4 (3.1)	+			5.3 (3.2)	+		
Pain assessment	–			9.8 (1.1)		+	9.9 (0.9)			+	9.9 (0.9)			+
Clinical depression screening	–			9.8 (1.1)		+	9.9 (0.7)			+	9.9 (0.8)			+
NHSN healthcare personnel influenza vaccination	–			9.8 (1.4)		+	9.6 (2.0)			+	9.7 (1.7)			+
Standardized hospitalization ratio	–			–			–				4.8 (3.1)	+		
Serum phosphorus	–			–			–				9.0 (1.5)			+
Ultrafiltration rate	–			–			–				4.1 (3.3)			+

ICH CAHPS = In-Center Hemodialysis Consumer Assessment of Healthcare Providers and Systems Survey; NA = not applicable; NHSN = National Healthcare Safety Network.

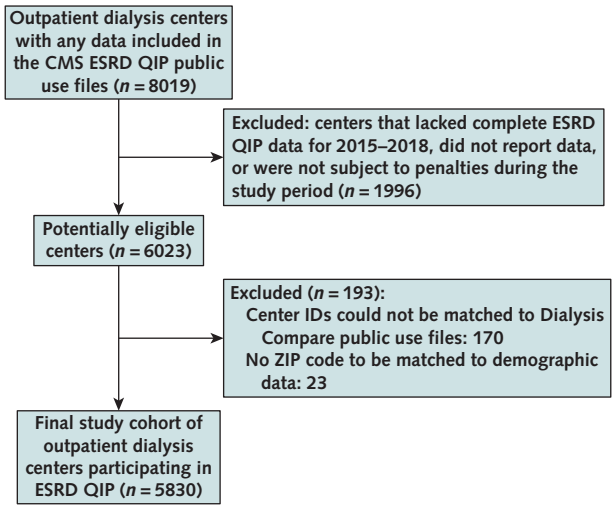
* A plus sign means that the measure was included in that year. A minus sign indicates that the measure was not included in that year. Technical specifications for each year are provided by the Centers for Medicare & Medicaid Service at www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/ESRDQIP.

Appendix Table 2. Number of Centers and Total Performance Score Ranges Included in Each Analysis*

Variable	2017 Analysis			2018 Analysis		
	Centers Below Threshold, *n*	Centers Above Threshold, *n*	Total Performance Score Range	Centers Below Threshold, *n*	Centers Above Threshold, *n*	Total Performance Score Range
All dialysis centers	640	1267	53–67	640	1267	53–67
All centers, risk-adjusted	639	1266	53–67	705	1474	52–68
Newly penalized centers	568	1177	53–67	568	1177	53–67
For-profit centers	575	1165	53–67	633	1353	52–68
Nonprofit centers	73	122	52–68	65	102	53–67
Chain-affiliated centers	577	1170	53–67	673	1546	51–69
Nonchain-affiliated centers	83	124	51–69	72	111	52–68
Large centers	162	282	54–66	215	420	52–68
Small centers	142	324	51–69	142	324	51–69
Centers in minority-predominant ZIP codes	165	295	54–66	193	374	53–67
Centers in nonminority-predominant ZIP codes	130	297	53–67	130	297	53–67
Centers in high-income ZIP codes	144	307	53–67	144	307	53–67
Centers in low-income ZIP codes	158	282	54–66	158	282	54–66

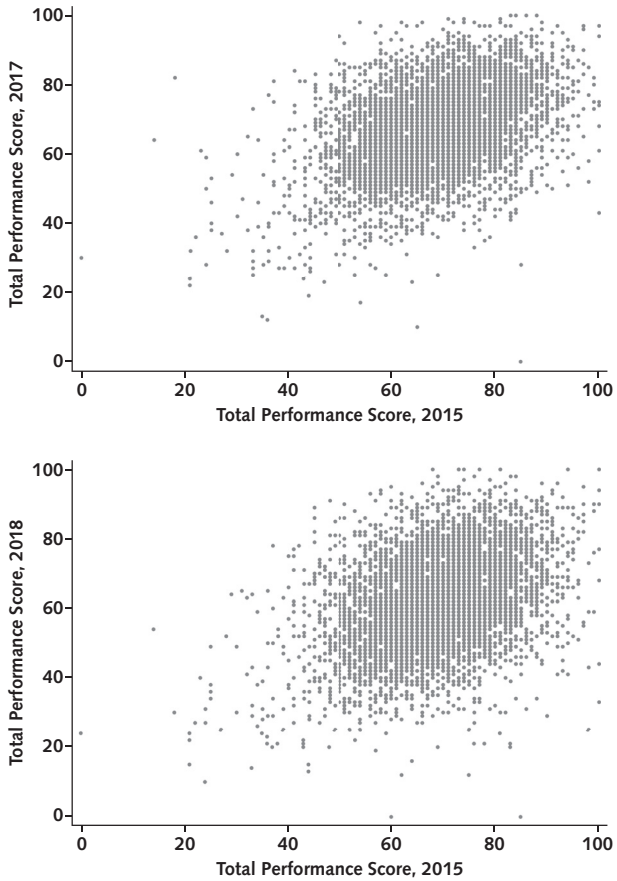
* A regression discontinuity analysis hinges on the assumption that centers immediately on either side of the threshold (penalization score of 60) are similar. Bandwidths represent the number of centers and the score range determined to be sufficiently close to be used to generate the regression discontinuity estimates. Bandwidths were derived by using data-driven selection in rdrobust (Stata). Centers were stratified by data contained in the Dialysis Facility Compare files maintained by the Centers for Medicare & Medicaid Services. Local demographic data were derived from the 2017 U.S. Census at the ZIP code level.

Appendix Figure 1. Study flow diagram.

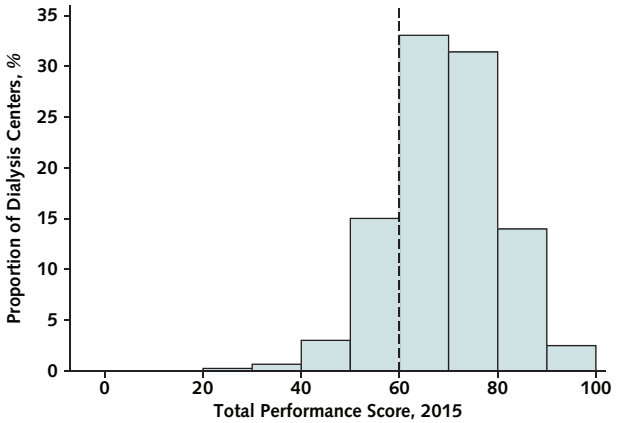


CMS = Centers for Medicare & Medicaid Services; ESRD QIP = End-Stage Renal Disease Quality Incentive Program.

Appendix Figure 2. Total performance scores in 2017 (*top*) and 2018 (*bottom*).



Appendix Figure 3. Distribution of total performance scores.



The vertical dashed line indicates the threshold for penalties. Centers with less than 60 points as their total performance score received financial penalties in 2015. Each bar represents the proportion of centers within 1 of 10 equally distributed bins between scores of 0 to 100.