Reputations and Firm Performance: Evidence from the Dialysis Industry

Subramaniam Ramanarayanan
Anderson School of Management, UCLA
Los Angeles, CA 90095
subbu@anderson.ucla.edu

Jason Snyder
Anderson School of Management, UCLA
Los Angeles, CA 90095
jason.snyder@anderson.ucla.edu

November 2011

Abstract:

We study the impact of information disclosure policies on firm performance by exploiting a natural experiment that quasi-randomly assigns reputations to firms based on their allocation to performance categories. Dialysis firm performance is publicly graded using three coarse performance categories based on patient survival rates: *better than expected, as expected,* and *worse than expected*. We exploit the underlying continuous performance measures used to create these categories to implement a regression discontinuity design. We find firms that are graded as performing *worse than expected* subsequently experience a reduction in patient mortality rates through improved patient care. Such firms also treat fewer informed patients. We do not find comparable effects for firms that are randomly assigned to the *better than expected* grade. The overall evidence is consistent with disappointing information being a significant motivator of firm behavior.

I. Introduction

Over the past fifty years, a large amount of literature in economics has highlighted the central role of information in decision-making.[1] One of the key insights derived from much of this work is that improving the quality and transparency of information leads to more desirable social outcomes.[2] Quality disclosure programs is one important area where these ideas have been applied. Numerous programs exist that provide consumers with information on firm performance and product quality in an effort to improve their decision-making. Restaurant hygiene grade cards (Jin and Leslie 2003), health-plan report cards (Scanlon *et al.* 2002; Jin and Sorenson 2006; Dafny and Dranove 2008), and hospital rankings (Pope 2009) are a few settings[3] where there has been significant academic research on the effects of quality disclosure on firm performance and consumer choice.

The empirical literature has come to two major conclusions regarding the impact of quality disclosure programs. First, at the consumer level, there is substantial evidence that quality disclosure leads to vertical sorting. All else equal, consumers are more likely to choose higher quality products post disclosure. Second, there is evidence that quality disclosure leads to subsequent improvements in product quality. For example, Jin and Leslie (2003) find a significant decline in hospital admissions for food borne illnesses after the adoption of restaurant hygiene grade cards in Los Angeles County.

Common to most of the previously studied disclosure programs is a quality score or an ordinal ranking[4] that is uniformly applied to all of the firms or products in a given marketplace. In practice, however, many disclosure programs are designed quite differently. Two distinct types of programs are frequently observed; those that celebrate

[1] Hayek (1945) and Stigler (1961) are early examples of research in this vein.
[2] The impact of information disclosure on social welfare is at times ambiguous; e.g., see Dranove, Kessler, McClellan, and Satterthwaite (2003).
[3] See Dranove and Jin (2010) for a comprehensive overview.
[4] For example, restaurant hygiene is graded on an A to F scale and hospitals (and universities) are ranked from 1 to 100.

the exceptional and those that shame the incompetent. Consider the firm quality disclosure programs of the Blue Ribbon Schools Program and Condé Nast Portfolio. The Blue Ribbon Schools Program "honors public and private elementary, middle, and high schools that are either high performing or have improved student achievement to high levels, especially among disadvantaged students."[5] Over the past twenty-five years, only around 6,000 schools have received a blue ribbon designation out of well over 100,000 eligible schools making this a program that recognizes the truly outstanding. In contrast to such a designation, Condé Nast Portfolio compiles a list of their "Toxic Ten",[6] a list of ten companies that are among the worst polluters in America and the ways in which they could be doing more for the environment.

There are several reasons to believe that focusing disclosure policies on the extremes of the firm quality distribution could be desirable. Tournament theory (Lazear and Rosen 1981) suggests that disclosing the identity of the very best firms can create a powerful incentive for all firms to improve product quality. On the other hand, the behavioral economics literature (Kahneman and Tversky 1979) provides compelling evidence suggesting that evaluations that fall short of expectations can be a markedly more powerful motivator compared to evaluations that exceed expectations.[7] This suggests that singling out the worst performers can provide a meaningful way to encourage improvements in product quality. Whether efforts should be focused on exposing the worst performers or praising the best is a question that is not convincingly addressed in the existing literature.

Establishing a causal relationship between these program features and future behavior poses a difficult challenge. The common methodology in the prior literature has

---

[5] See the department of education website: <http://www2.ed.gov/programs/nclbbrs/index.html>

[6] From Condé Nast Portfolio's website: <http://www.portfolio.com/news-markets/national-news/portfolio/2008/02/19/10-Worst-Corporate-Polluters/>

[7] An example of this comes from Mas (2006) who finds that arrest rates fall when police officers receive pay raises that fall short of their expectations. A smaller sized effect is seen when police officers receive raises that exceed expectations. There is limited evidence of the role of expectations in shaping behavior in product disclosure markets. Dranove and Sefkas (2008) find that vertical sorting is dependent upon the extent to which the information that is being disclosed actually updates firm and consumer beliefs based on their prior expectations. If the information being disclosed is not surprising, it is less likely that behavior will change.

been to use a differences-in-differences design to assess outcomes before and after the implementation of the disclosure program relative to a control group.[8] A particularly troubling possibility is that these methods can attribute the impact of the program to mean reversion. When a regulator observes distressing signals in a marketplace there could be a tendency to pursue information disclosure policies as a remedy. For example, a health plan disclosure law could be enacted because lawmakers receive numerous complaints about the complexity of health insurance. If subsequent improvements in the marketplace are seen, it is possible that the results could be due to mean reversion. Quality is naturally variable over time and it is possible that after the disclosure program is implemented there will be a return to the long-term average that would have occurred in the absence of the disclosure program. Mean reversion can be exacerbated when looking at the extremes of the quality distribution where difference in difference designs can substantially overstate the impact of disclosure programs.[9]

To overcome these challenges, we use a natural experiment from the dialysis industry that quasi-randomly assigns reputation to firms based on their assignment to different coarse performance categories. The most prevalent modality of treatment for kidney failure is in-clinic dialysis treatments where the patient visits the dialysis center three times a week to undergo treatment. Approximately 20% of patients die while on dialysis each year and there is considerable dispersion in quality across facilities. [10] Up until late 2010, the only available measures of firm quality available to patients seeking to undergo dialysis were three coarse performance grades based on patient survival rates: *as expected*, *better than expected*, and *worse than expected*.[11] Because of the coarseness of the categorical ratings, facilities with very similar patient survival rates could receive

---

[8] Pope (2009) is a notable exception. He uses an instrumental variables approach to assess the impact of hospital rankings on future admissions.

[9] These problems are prominent in other domains, Chay, et. al (2005) find that difference in difference designs can seriously inflate the estimated impact programs relative to a quasi experimental design because of the above rationales.

[10]http://www.theatlantic.com/magazine/archive/2010/12/-8220-god-help-you-you-39-re-on-dialysis-8221/8308/

[11] In late 2010, data on actual risk adjusted mortality rates were released for the years 2002-2010. Prior to that year the National Kidney Foundation was extremely reluctant to disclose the underlying risk adjusted mortality data and consumers were privy only to the information conveyed by the coarse performance categories.

very different grades. For example, a facility that had a 90th percentile national ranking based on patient mortality could receive an *as expected* performance grade while a facility at the 91st percentile could be assigned a *worse than expected* grade, [12] despite these facilities being essentially of the same quality. Prior to the categorical quality disclosure, both facilities should have had essentially similar characteristics. The only difference is that the facility with the slightly higher mortality percentile rank received a much lower categorical rating. Using recently available archival performance data, we use a regression discontinuity design to isolate the causal impact of changes in the information disclosed about the facilities on future performance.

We find that a facility that receives a *worse than expected* performance grade experiences a substantial improvement in future performance in comparison to a facility that is graded *as expected*. We provide evidence consistent with these improvements in patient survival being caused by the effect of information disclosure as opposed to the facility altering its patient profile to manipulate its ranking. We show that patients who are more informed sort away from facilities that are ranked *worse than expected* but find no evidence to suggest vertical sorting of the nature seen in previous studies. In other words, facilities that are rated *worse than expected* see no overall change in patient volumes but do treat a higher percentage of patients who were likely unaware of the ratings.

In contrast, we find that distinguishing facilities between those that performed *better than expected* and *as expected* has little impact on relevant outcomes. Taken together, we find strong causal evidence that disclosure policies impact firm performance; however, the results are heavily skewed towards the low end of the firm quality distribution.

The paper proceeds as follows: Section II describes the institutional setting and the data in greater depth, Section III outlines the regression discontinuity design used in

---

[12] Note that lower percentiles represent better performance here since the performance measure is based off patient mortality. We expand on the performance measure in Section II.

the empirical estimation, Section IV discusses the results, and Section V offers concluding observations.

II. Institutional Setting and Data

A. End Stage Renal Disease

End Stage Renal Disease (ESRD) refers to a stage of Chronic Kidney Disease (CKD) when the kidneys completely fail in their function of removing waste from the body. Once the patient's condition has deteriorated to this stage, the only options available for treatment are dialysis or transplantation.[13] Most physicians view transplantation as the preferred treatment primarily because ESRD patients that undergo an organ transplant live longer and healthier lives in comparison to patients treated with dialysis. However, the number of healthy organs available for transplantation is heavily outnumbered by the number of patients suffering from ESRD leading to a major kidney shortage.[14] As a result, nearly 70 percent of ESRD patients in the US (approximately 400,000 patients) currently undergo dialysis every year as treatment for kidney failure.[15]

In 1972, the Social Security Act extended Medicare Part A and Part B benefits to individuals with ESRD regardless of age (Nissenson and Rettig 1999). This entitlement currently covers over 90 percent of all patients suffering from ESRD in the United States. Medicare covers both inpatient (under Part A) and outpatient (under Part B) dialysis treatments and typically pays 80 percent of the approved amount with the patient being responsible for the remaining 20 percent. Patients may pay for this coinsurance out-of-pocket or through supplemental insurance policies such as Medicaid or Medigap.

---

[13] These treatments are not exclusive. Many patients undergo dialysis while on the waitlist for a kidney transplant.

[14] To illustrate the extent of shortage, consider the following. According to the National Kidney Foundation, over 80,000 patients were on the waitlist to receive a kidney transplant in 2009. This compares to a total of 16,500 kidney transplants performed in the U.S. in 2008.

[15] http://kidney.niddk.nih.gov/kudiseases/pubs/kustats/

Since 1983, Medicare has reimbursed dialysis facilities a fixed fee for each treatment. This payment is broken up into a base rate which is intended to cover provider costs, and covers the entire bundle of services, tests and certain drugs for up to three dialysis sessions per week.[16] This base rate is then adjusted to account for differences in patient case mix based on patient age and Body Mass Index.[17] Finally, differences in local input prices (i.e. wages) across facilities are also incorporated into the final payment.[18] An important implication of this near universal coverage is that dialysis facilities compete on quality given that prices are mostly fixed. ESRD accounted for approximately 6% of the total Medicare budget in 2010 (Fields 2010). Given that Medicare expenses accounted for approximately 13% of the 2010 federal budget, *ESRD expenditures made up almost 1% of the entire federal budget in 2010.*[19]

B. The Dialysis Industry

Dialysis is a treatment that is designed to replicate the cleaning function of kidneys when they fail. Dialysis treatments help ESRD patients live longer but are not intended as a permanent cure for kidney failure. Dialysis treatment falls into one of two kinds each of which takes a different approach to removing waste from the bloodstream. Hemodialysis uses a special membrane to filter the blood and is usually performed at a dialysis facility. Peritoneal dialysis uses the lining of the abdominal cavity, the Peritoneum, to filter the blood and is usually performed at the patient's residence. Patients may choose to switch from one mode of dialysis to the other as their treatment progresses.

The vast majority of dialysis patients in the U.S. are treated in one of the approximately 5,000 nationwide dialysis centers three times a week, with each treatment lasting for three to five hours. Over 90 percent of these centers are freestanding facilities

[16] For 2012, the base rate is $234 for freestanding and hospital-based facilities

[17] Starting 2011, the patient level adjustment will also account for six other comorbities. Source: Medicare Payment Advisory Commission, accessible at http://www.medpac.gov

[18] Starting 2011, Medicare is phasing in a new Prospective Payment System which bundles together all dialysis services and items that were previously billed separately. This change occurs outside the timeframe of our data and hence does not affect our analyses.

[19] Expenditures incurred by patients with a diagnosis of kidney disease made up 31 percent of Medicare expenditures in 2009 (source: http://www/usdrs.org)

and in addition to dialysis services, may provide lab testing and drug infusion services. A typical center provides around 50 treatments a day using 15-20 dialysis stations. Each center is required to have a medical director who must be board-certified in internal medicine or pediatrics and have experience in dealing with ESRD patients (Lawler et al. 2003). In addition, the Center for Medicare and Medicaid Services (CMS) mandates the presence of at least one licensed registered nurse, a social worker and a dietitian. Centers may employ additional patient care technicians, but at least one licensed health care provider (such as a doctor or a registered nurse) needs to be present at the center when a patient is undergoing dialysis. Staffing ratios vary by state, and few states have regulations regarding these numbers (Wolfe 2011).[20] Mortality rates on dialysis are grim. Approximately 20% of patients die within their first year on dialysis and 65% die within 5 years.[21]

The market structure of the dialysis industry has undergone dramatic changes over the last decade. While the number of dialysis facilities has grown from around 2000 units in 1991 to over 5000 units in 2009, the industry has also become increasingly concentrated over time; the two largest dialysis providers, Davita and Fresenius, together accounted for over 60 percent of market share in 2009 (USRDS 2011). [22] Nearly 80 percent of dialysis facilities are designated as being under for-profit ownership (Fields 2010).

C. The Dialysis Facility Compare Data

As noted earlier, mortality rates for patients undergoing dialysis are quite high. There is, however, considerable dispersion in mortality across dialysis facilities. In response to this variation in quality, Medicare released the Dialysis Facility Compare (DFC) tool in 2000. The primary impetus for this program came from the Balanced Budget Act of 1997

[20] As an example, Georgia mandates a staffing ratio of Registered Nurses to Dialysis patients of 1:10, while Texas requires a ratio of 1:12. In addition, the National Kidney Foundation releases recommended staffing ratios in terms of the number of dietitians (1:100) and social workers (1:75) per patient undergoing dialysis. See Wolfe (2011) for more details.

[21] Mortality rates for patients undergoing dialysis are similar to patients having stage III colon cancer.

[22] Independently owned facilities accounted for 15 percent with hospital based facilities and other smaller chains accounting for the rest.

which required the Center for Medicare and Medicaid services (CMS) to *"develop and implement (by January 1, 2000) a method to measure and report the quality of renal dialysis services provided under the Medicare program"* (Frederick *et al*. 2002). DFC was first introduced on the www.medicare.gov website on January 19, 2001 and provided consumers with information on the location, hours, and quality (as measured by the Standardized Mortality Ratio) of almost all of the nation's dialysis facilities. The Standardized Mortality Ratio (SMR) compares the observed mortality rate in a particular facility to the death rate that would be expected based on national death rates for patients with characteristics similar to those treated at the facility. The SMR is typically adjusted for patient demographics such as age, sex, race and ethnicity and comorbidities such as diabetes, BMI, duration of ESRD, as well as regional variables such as state population death rates. Based on recommendations from a Consumer Information Workgroup,[23] CMS decided not to report patient survival rates (as measured by facility SMRs) in a continuous manner but to report it in three categories based on specific cut points: *as expected*, *better than expected*, and *worse than expected*. The National Kidney Foundation uses a four-year window as the basis for computing these SMR categories. For example facility mortality between the years 2002-2005 would be used as a basis for constructing the categories that would be released in the middle of 2006.

In December 2010, Propublica.org, an independent, non-profit investigative news outlet, made available to the public the precise mortality data for all four-year windows between 2002 and 2009. Robin Fields, an investigative reporter and senior editor with Propublica.org, obtained this information through filing multiple Freedom of Information Act requests over the course of two years. She made this information publicly available on the Propublica.org website out of concern that the coarse SMR categories (reported by DFC on the CMS website) were not sufficient to adequately compare facilities on quality.[24]

---

[23] This workgroup included representatives of physicians, nurses, patients and social workers, and facility administrators (Frederick *et al.* 2002).

[24]Fields (2010) effectively describes the inadequacy of the current DFC ratings using an example: "*Innovative Renal Care and Midtown Kidney Center, clinics about two miles apart in Houston, had similar stats on Dialysis Facility Compare in 2007, including "as expected" survival rates. But the full data show that Innovative Renal's average annual death rate—after factoring in patient demographics and*

We construct a dataset containing facility level SMRs (raw scores, as well as the coarse categories reported by CMS) and as a number of facility characteristics for the years 2008-2009. Given that mortality rates are computed using four-year windows, the dataset contains information on patient survival rates in dialysis facilities starting with the 2004-2007 timeframe. The sample includes information on all 4,665 firms that received performance evaluations on the *Dialysis Facility Compare* website in December of 2008. We are able to link these firms to their past and future mortality rates, organizational form, ownership information, and patient characteristics on a yearly basis from 2004-2009. All of our analysis is performed at the facility/year level.

Figure 1 shows the distribution of SMRs for dialysis facilities in the U.S. constructed using data from the years 2004-2007. The wide dispersion in patient survival rates referenced earlier can be readily seen here. At the bottom 10th percentile of the distribution a facility has a 30% lower than expected mortality rate. At the top 90th percentile a facility has a 33% higher than expected mortality rate.[25] Prior research has investigated some of the determinants of these mortality differences across centers. Garg *et al.* (1999) find that for-profit ownership of dialysis facilities is associated with higher mortality rates and Powe *et al.* (2002) report that dialysis administrators believe that patient education, staffing ratios, and level of wages are crucial determinants of facility quality. In a cross-sectional study of 90 dialysis facilities, Spiegel *et al.* (2010) find that dialysis centers that were categorized as performing *better than expected* were associated with more engaged patients, and better communication and coordination between physicians and staff.

When the program was introduced in 2001, facilities were categorized into one of three categories based on patient survival rates: *better than expected (by 20 percent or more), as expected, and worse than expected (by 20 percent or more).* Specifically, a

---

complicating conditions—was 34 percent higher than expected. Midtown's average rate was 15 percent lower than expected. Dialysis Facility Compare has since changed Innovative's survival rating to "worse than expected," but how much worse? The unpublished 2009 data reveal that the clinic performed more poorly, versus expectations, than 92 percent of all facilities nationwide."

[25] Because approximately 20% of patients die each year, these differences are large.

facility was categorized as having a patient survival rate that was *better (worse) than expected* if the upper (lower) confidence limit for the facility's SMR is less (greater) than 0.8 (1.2). This categorization led to the vast majority of facilities (96 percent) being designated as belonging to the *as expected* category prior to 2008. In 2008, CMS updated the thresholds based on which facilities were assigned to different performance categories in order to "help consumers make better distinctions among facilities' survival rates".[26] In particular, facilities were now classified as performing *better than expected* if the facility SMR was less than 1.00 and statistically significant ($p<0.05$). If the facility SMR is greater than 1.00 and statistically significant ($p<0.05$), the facility was now classified as performing *worse than expected.* All other facilities were classified as performing *as expected*.[27]

As a direct result of this change, far more firms were now classified as performing better or worse than expected; looking at the SMRs computed using the 2004-2007 four-year window, approximately 80% of facilities now received an *as expected* grade. **Figure 2 and Figure 3** illustrate the impact of the change in thresholds on the classification of dialysis facilities into performance categories using data from the 2004-2007 window. We study the impact of the grades coming from the 2004-2007 four-year window because prior to then, there were too few observations in the *better than expected* and *worse than expected* categories to effectively use the proposed regression discontinuity design.

Table 1 presents summary statistics for the 4665 dialysis facilities in our sample. A majority of these facilities (80 percent) are associated with for-profit ownership and are affiliated with a national or regional chain. Each facility admits around 22 new patients on average each year, with nearly 70 percent of these patients being referred to the

[26] In addition to the change in the way in which dialysis facilities were grouped into patient survival categories, Dialysis Facility Compare was also modified to report two anemia measures – the percentage of patients whose hemoglobin was considered too low (below 10g/dL) or too high (above 12 g/dL) – in contrast to earlier versions of the tool which only reported the proportion of patients with high hemoglobin levels. This modification was undertaken based on new guidelines imposed by the Food and Drug Administration. See CMS press release dated November 20, 2008 titled "Medicare Publishes New Information on Quality of Care at Dialysis Facilities" for more details.
[27] Facilities with three or fewer deaths are not included in the classification.

facility by a nephrologist. We also note some discrepancy in the number of observations across variables. This occurs for two reasons. First, 153 of the 4,665 facilities reported on the December 2008 dialysis facility compare website closed in 2009. This has the potential to introduce survivorship bias into our results. In unreported results, we find that firm survival is not related to the coarse categorical grades which helps mitigate this concern. Second, some of the reported variables are not uniformly recorded across all firms. For example, we observe 4,512 firms with data in 2009 but only 4,464 with information on whether a new patient visited a nephrologist in the prior year. These missing observations are a relatively small portion of the overall data and do not appear to have any systematic correlation with firm characteristics or performance.

III. Identifying the Causal Impact of Information Disclosure on Performance

Our primary methodological approach is to use a regression discontinuity design to estimate the causal effect of information disclosure on future performance outcomes. Before discussing the specifics of the empirical approach, we outline the timeline of the various events relevant to the study. Understanding the timeline is crucial to following the research design of the study.

A. Timeline of Information Disclosure

As discussed in Section II, the facility reports using data from the 2004-2007 four-year window were the first to classify firms into performance categories using the new thresholds. This information was revealed at different times to the facilities, consumers, and researchers.

- *June 2008*: Each facility receives a Dialysis Facility Report (DFR) from CMS that enable them to learn which coarse performance category it is assigned to, based on the facility SMRs computed using the 2004-2007 four-year window. Given that the DFR contains the actual value of the SMR, each facility learns how close it is to the threshold.[28] Facility level performance data from 2008 would be the first year to show a response to this information.

- *December 2008*: The *Dialysis Facility Compare* website is updated with information on each facility's performance using the 2004-2007 four-year window. Consumers learn what coarse performance categories all facilities fall into. Facilities also learn about the standing of their competitors with respect to these coarse performance categories. Facility level consumer choice data from 2009 would be the first year to show a substantial response to this information. Henceforth in this paper, we will use *as expected*, *better than expected*, and *worse than expected* to refer to the performance categories being generated using the mortality data from the 2004-2007 four-year window.

- *December 2010*: The precise mortality data from all four-year windows between 2002 and 2009 is made available on Propublica.org and accessible to the general public as well as researchers. Prior to this date, the only information on facility performance available to the public was the coarse performance categories on the *Dialysis Facility Compare* website (Fields 2010).

B. The Regression Discontinuity Design

According to Imbens and Lemieux (2007), the idea behind using the regression discontinuity designs for evaluating causal effects of interventions is that "…assignment to a treatment is determined at least partly by the value of an observed covariate lying on either side of a fixed threshold."[29] The fundamental identifying assumption is that close to a threshold of interest all other characteristics and choices that could influence an

[28] Note that the SMR is based on the comparison of each facility's performance to performance of firms nationwide, so each facility is unable to gauge how close they are to the performance thresholds based on their performance alone.
[29] Note that the covariate may itself be associated with the outcome, but the key assumption is that this association is smooth. Therefore, any discontinuity of the outcome as a function of the covariate at the threshold is taken as evidence for a causal effect of the treatment. See Imbens and Lemieux (2007) for further discussion and for examples of various settings in which the regression discontinuity approach has been employed to estimate causal effects of treatments.

outcome will be orthogonal to the treatment being studied. Regression discontinuity design is a well-established methodology with a straightforward causal interpretation.

Figure 4 illustrates the application of the regression discontinuity design in this context. We plot the relationship between subsequent performance of each facility (in 2008) on the y-axis and the 95th percent lower confidence interval for the 2004-2007 SMR on the x-axis. In this figure, we measure the performance of each facility in 2008 on the y-axis as the national percentile ranking of the facility based solely on their SMR in 2008.[30] Lower values along this axis would therefore point to higher levels of performance. For example a facility with a percentile ranking of 1 would be in the top 1% of all facilities in terms of patient survival based on the 2008 mortality data. On the x-axis, a facility that has an SMR with a lower confidence interval just above 1 is classified as performing *worse than expected* while a facility that has a lower confidence interval just below 1 receives an *as expected* grade.

The figure provides compelling evidence of a discontinuity precisely at the point where a firm moves from the *as expected* category to the *worse than expected* category. A firm that just barely falls in the *worse than expected* category is seen to be much more likely to improve subsequent performance relative to a firm that just barely falls into the *as expected* category. Visually it appears unlikely that the change in the 2008 SMR percentile ranking is caused by any factor other than having been just barely rated as performing *worse than expected*. In our empirical specifications, we aim to estimate the magnitude of this change for various outcome measures when firms receive information about their relative standing with respect to performance. In our analyses, we separately estimate the difference between firms being assigned to the *as expected* and *worse than expected* categories from the difference between being assigned to the *as expected* and *better than expected* categories. Our principal specification is the following equation estimated on facility-level data, where *i* indexes a firm:

$$Outcome_i = \alpha + \beta * Threshold_i + \gamma * N^{th} \text{ degree polynomial of } CI_i + \varepsilon_i \qquad (1)$$

[30] This percentile ranking only uses mortality data from 2008

Threshold is an indicator variable equal to 1 if the firm is assigned to a *better than expected* or *worse than expected* category.[31] The coefficient estimate on *Threshold* is the size of the jump being estimated. **Figure 5** provides a graphical representation of what this estimation process yields when applied to figure 4. The coefficient estimate on *Threshold* is the difference between the two estimated lines when the lower confidence interval equals 1. The figure also clearly illustrates the role of the polynomial control variables. The polynomials control for the underlying trend and allow *Threshold* to identify the discontinuous break in the data. The greater the number of degrees on the polynomial controls, the better the fit to the underlying trend.

IV. Results

A. Testing Validity of the Regression Discontinuity Design

As discussed earlier, the key identifying assumption of the regression discontinuity design is that all potential confounding variables are orthogonal to the treatment. While this assumption cannot ever be comprehensively tested, it is possible to provide evidence that is consistent with the assumption. One approach is to show that potentially confounding covariates are not systematically distributed on either side of the threshold, i.e. "testing for the null hypothesis of a zero average effect on pseudo outcomes that are known not to be affected by the treatment" (Imbens and Lemieux, 2007). We implement this test and report the results in **Table 2** and **Table 3**. In each table, we include polynomials of degree 3 as controls, and examine a bin of width 0.2 (i.e., width of 0.1 on each side around 1).

The coefficient estimates in Table 2 indicate that there is no evidence of a discontinuity in any of the 2007 outcome variables in response to barely exceeding the

[31] In none of our specifications will *worse than expected* and *better than expected* categories be included in the same regression equation.

worse than expected threshold.[32] This implies that there is no evidence of strategic sorting by firms around the threshold. We reach similar conclusions when we examine the *better than expected* threshold in Table 3. The sole exception is the estimate in Column 4 which seems to suggest that firms that were part of a chain in 2007 are more likely to be just barely inside the *better than expected* threshold. However, this result only holds at the 10% confidence level. It is to be expected that with 14 different variables, one or more of them is significant at the 10% confidence level. Importantly column (1) of tables 2 & 3 shows that the standardized mortality ratio from 2004-2007 is not related to the thresholds.[33] This is a precisely estimated zero. Taken together there is little evidence to suggest that the treatment violates random assignment.

B. *The Impact of Disclosure Policies on Performance*

Table 4 presents coefficient estimates of our main specification, i.e. impact of being assigned to the *worse (or better) than expected* performance categories on subsequent outcomes. The impact of being assigned to the *worse than expected* and *better than expected* performance categories are presented in the first and last four columns, respectively. The results in the first four columns are consistent with the pattern seen earlier in Figure 5. In Column 1, we include a 3rd degree polynomial of the lower confidence interval as a control, and a sample window of all observations with lower confidence intervals between 0.9 and 1.1.[34] The magnitude of the effect is quite striking: a facility that is just barely categorized as having patient survival rates that are *worse than expected* experiences a 14 percentile improvement in the 2008 SMR percentile ranking.

In columns 2 and 3, we show that this result is robust to two major factors that influence regression discontinuity estimates: the size of the sample window and the

[32] This is not surprising since the coarse performance categories from the 2004-2007 four-year window were not known until 2008.
[33] Since the thresholds come from the confidence intervals there is no reason to believe this is necessarily true or false.
[34] This corresponds to a sample window range of .2. For *better than expected* threshold regressions, a sample window range of 0.2 would imply that the upper confidence interval would fall between 0.9 and 1.1.

degree of the control polynomial. In column 2, we use a 10th degree polynomial as a control as opposed to a 3rd degree polynomial and find similar results. In column 3, we expand the sample window to include all facilities with a lower confidence interval between 0.75 and 1.25, i.e. a sample window of width 0.5. This increases the number of firms in the estimation sample from 590 to 1823. While the coefficient drops in magnitude, the economic significance is still substantial. In column 4, we include controls for for-profit status, whether a chain owned the facility, total number of dialysis stations, and the standardized mortality ratio from 2004-2007. The coefficient on the indicator is effectively unchanged. This suggests that the treatment is randomly assigned.

In columns 5-8, we present results from similar specifications that examine the 2008 SMR percentile ranking of firms who just barely pass the *better than expected* threshold. The coefficients are much smaller in magnitude and none of them are statistically significant. Taken together, the results in Table 4 suggest that poorly performing firms respond to negative disclosure by improving quality. The response of exceptional firms to positive disclosure is limited and difficult to distinguish from zero. In addition, the quality improvement is economically significant

C. Testing for Mechanisms Underlying Performance Improvement

How does this performance improvement occur? An important component of high quality dialysis care is how well a patient's anemia is managed. Most patients on dialysis do not have an ideal red blood cell count: they are anemic. Improper management of anemia endangers patient well-being and increases mortality risk. A blood test administered to patients undergoing dialysis that checks for hemoglobin levels provides a measure of the extent to which anemia is being properly managed by the facility. Ideal hemoglobin levels are between 10 g/dL and 12 g/dL. While all centers are required to frequently check and report hemoglobin levels, they are not required to have all of their patients at this ideal level. To achieve such an optimal level of hemoglobin, a facility must actively adjust anemia medication.

One big hurdle to achieving ideal hemoglobin levels is that the facilities' incentives are not well aligned. Since Medicare reimburses anemia medication, a facility can over-prescribe such drugs to patients to increase revenues. When excess medication is used, hemoglobin levels can fall below 10 g/dL. Well-managed anemia means that facilities are actively engaged with patients in medication management and avoiding the revenue related temptations to over-prescribe. The Dialysis Facility Reports include data on the percentage of patients treated at a facility whose anemia is well managed.[35] We use this data to construct two variables: whether a facility is above the median in anemia management and whether the facility is in the top 10 percent of firms based on anemia management in the US during 2008.

Table 5 presents estimates from specifications where we use these variables as dependent variables in a regression discontinuity model analogous to the one estimated in Table 4. In column 1 of Table 5, we find no evidence that being just barely within the *worse than expected* threshold leads to above median anemia management.[36] However, in column 2, we find that the probability that a facility is among the top 10 percent of firms nationwide in anemia management increases substantially if the facility falls just inside the *worse than expected* threshold. This finding is robust to changing the sample window range and the degree of the polynomial control variables in column 3. This suggests that the improvement in care is not uniform across treated firms. A limited number of firms exhibit dramatic performance improvement. In contrast, we find no such improvement for facilities that are just about classified as having *better than expected* patient survival rates (columns 4-6).

Table 6 presents results from specifications that aim to estimate the impact of information disclosure on the number and type of new patients choosing a facility. Because patients do not see the coarse performance data until December of 2008, we would expect the impact of the new disclosure policy to be felt on the number or

[35] Unlike the mortality data, this measure is not risk adjusted. These results should therefore be interpreted with caution.

[36] We use linear probability models with these discrete dependent variables because regression discontinuity designs quite often lead non-linear maximum likelihood estimators such as probits or logits to not converge.

composition of new patients only in 2009. Column 1 shows the impact on the number of new patients at the facility in 2009. The result is an imprecisely estimated zero effect. From column (1) it could be the case that just barely being in the *worse than expected performance* category has no effect on the number of new patients. It is also possible that it could reduce the number of new patients by 25%, which would be an economically significant effect.[37]

The summary statistics in Table 1 indicate that approximately 30 percent of patients did not visit a nephrologist prior to starting dialysis. Nephrologists are kidney specialists who help patients manage the course of ESRD. Patients benefit from visiting a nephrologist as such a specialist is more likely to be aware of the *Dialysis Facility Compare* website and would also be knowledgeable about the quality of local dialysis facilities when referring a patient to a dialysis center. Our assumption is that if a patient did not see a nephrologist before starting dialysis they are less likely to be aware of the facility performance rankings. In columns 2 and 3 of Table 6, we find strong evidence that when a facility is just barely within the *worse than expected* threshold, an additional 10 percent of the new patients in 2009 will not have seen a nephrologist prior to dialysis. This suggests that informed patients are shying away from facilities that are assigned as performing *worse than average* on mortality. In columns 4-6, we find no impact of just barely being within the *better than average* threshold on the number of new patients or the percentage of new patients that are relatively less informed.

V. Conclusions

The evidence in this paper is consistent with information disclosure programs being most effective when the incompetent are shamed. One important caveat when trying to generalize this result beyond the dialysis industry is that firms in the dialysis industry are capacity constrained. The best quality firms can only treat a limited number of patients.

[37] This number comes from the fact that for the parameter and standard error the 95% lower confidence interval is approximately equal to -5. Since the average firm takes on 20 new patients a year this is an approximately 25% reduction.

In industries where firms are not capacity constrained, it is likely that information disclosure programs that praise the exceptional are as effective if not more so.

Our results showing a lack of vertical sorting among consumers stands in contrast to the findings in the existing literature. It raises the possibility that mean reversion may be driving the results in the existing literature and points to the need for implementing more compelling research designs in the information disclosure literature to adjudicate this debate. Combining the above result with the result on nephrologist visits points to a practical implication for the design of the *Dialysis Facility Compare* program. These results suggest that the information on the website is not being utilized by all of the patients seeking dialysis treatment. Finding a means to make the information more salient to consumers could lead to important welfare improvements.

VI. Bibliography

Chay, K., P. McEwan and M. Urquiola. 2005. "The central role of noise in evaluating interventions that use test scores to rank schools", *American Economic Review*, 95, 1237-58

Dafny, L, and D Dranove. 2008. "Do report cards tell consumers anything they don't already know? The case of Medicare HMOs." *The RAND Journal of Economics,* 39(3): 790–821.

Dranove, D, Kessler, D., McClellan, M. and M. Satterthwaite et al. 2003. "Is More Information Better? The Effects of 'Report Cards' on Health Care Providers." *Journal of Political Economy*, 111(3): 555–588.

Dranove, D, and A Sfekas. 2008. "Start spreading the news: A structural estimate of the effects of New York hospital report cards",, *Journal of Health Economics,* 27(1): 1201–1207.

Dranove, D, and G Jin. 2010. "Quality disclosure and certification: theory and practice." *Journal of Economic Literature*, 48(4): 935–963.

Frederick, P., N. Maxey, S. Clauser, and J. Sugarman. 2002. "Developing Dialysis Facility-Specific Performance Measures for Public Reporting", *Health Care Financing Review*, 23(4), 37-50.

Fields, R. 2010. "God Help You. You're on Dialysis." *Atlantic Monthly*: 1–11.

Garg, P, K Frick, M Diener-West, and N Powe. 1999. "Effect of the ownership of dialysis facilities on patients' survival and referral for transplantation", *New England Journal of Medicine*, 341(22): 1653–1660.

Imbens, G, and T Lemieux. 2007. "Regression discontinuity designs: A guide to practice", *Journal of Econometrics*, 142(2): 615–635.

Jin, G, and A Sorenson. 2006. "Information and consumer choice: The value of publicized health plan ratings", *Journal of Health Economics*, 25(2): 248–275.

Jin, G, and P Leslie. 2003. "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards", *Quarterly Journal of Economics*, 118(2): 409–451.

Kahneman, D, and A Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk", *Econometrica*, 47(2): 263–291.

Lawler, M., G. Doeksen, F. Eilrich and V. Schott. 2003. "A Systems Development Guide for a Kidney Dialysis Centre", available online at <http://www.ruralhealthworks.org/dl_additional.html>

Lazear, E, and S Rosen. 1981. "Rank Order Tournaments as Optimum Labor Contracts", *Journal of Political Economy,* 89(5): 841–864.

Mas, A. 2006. "Pay, reference points, and police performance", *Quarterly Journal of Economics*, 121(3): 783–821.

Nissenson, A, and R Rettig. 1999. "Medicare's end-stage renal disease program: current status and future prospects", *Health Affairs,* 18(1): 161–179.

Pope, D. 2009. "Reacting to rankings: Evidence from", *Journal of Health Economics*, 28(1): 1154–1165.

Powe, N, M Thamer, W Hwang, N Fink, E Bass, J Sadler and N Levin. 2002. "Cost-quality trade-offs in dialysis care: A national survey of dialysis facility administrators", *American Journal of Kidney Diseases*, 39(1): 116–126.

Scanlon, D, M Chernew, and C McLaughlin. 2002. "The impact of health plan report cards on managed care enrollment", *Journal of Health Economics,* 21(1): 19–41.

Spiegel *et al*. 2010. "Dialysis Practices that Distinguish Facilities with Below- versus Above- Expected Mortality", *Clinical Journal of the American Society of Nephrology*, 5 (11), 2024-33.

USRDS 2011. "United States Renal Data System Annual Report", available online at < http://www.usrds.org/adr.aspx>

Wolfe, W. 2011. "Adequacy of Dialysis Clinic Staffing and Quality of Care: A Review of Evidence and Areas of Needed Research", *American Journal of Kidney Disease*, 58(2), 166-176.

Zhang, Y., D. Cotter and M. Thamer. 2011. "The Effect of Dialysis Chains on Mortality among Patients Receiving Hemodialysis", *Health Services Research*, 46:3, 747-67.

Figure 1: Kernel density of
standardized mortality ratio 2004-2007

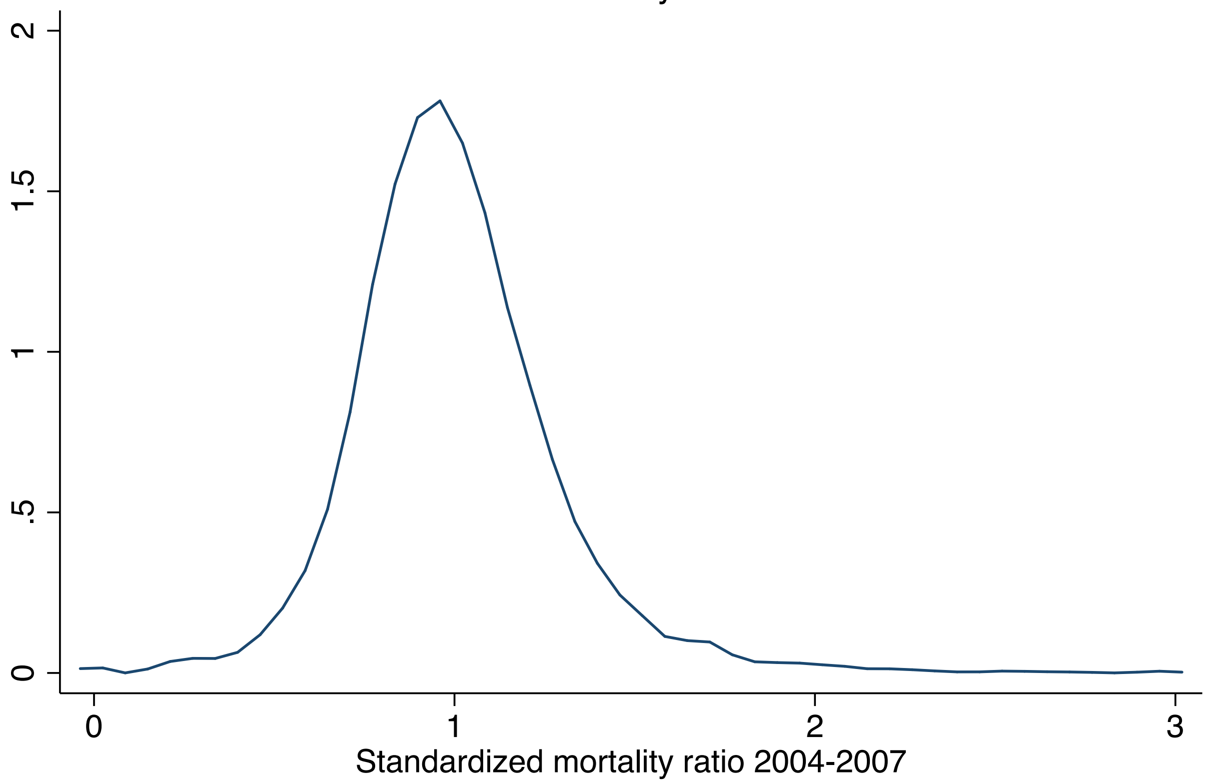


Figure 2: Kernel density of lower confidence interval of the 2004-2007 standardized mortality ratio

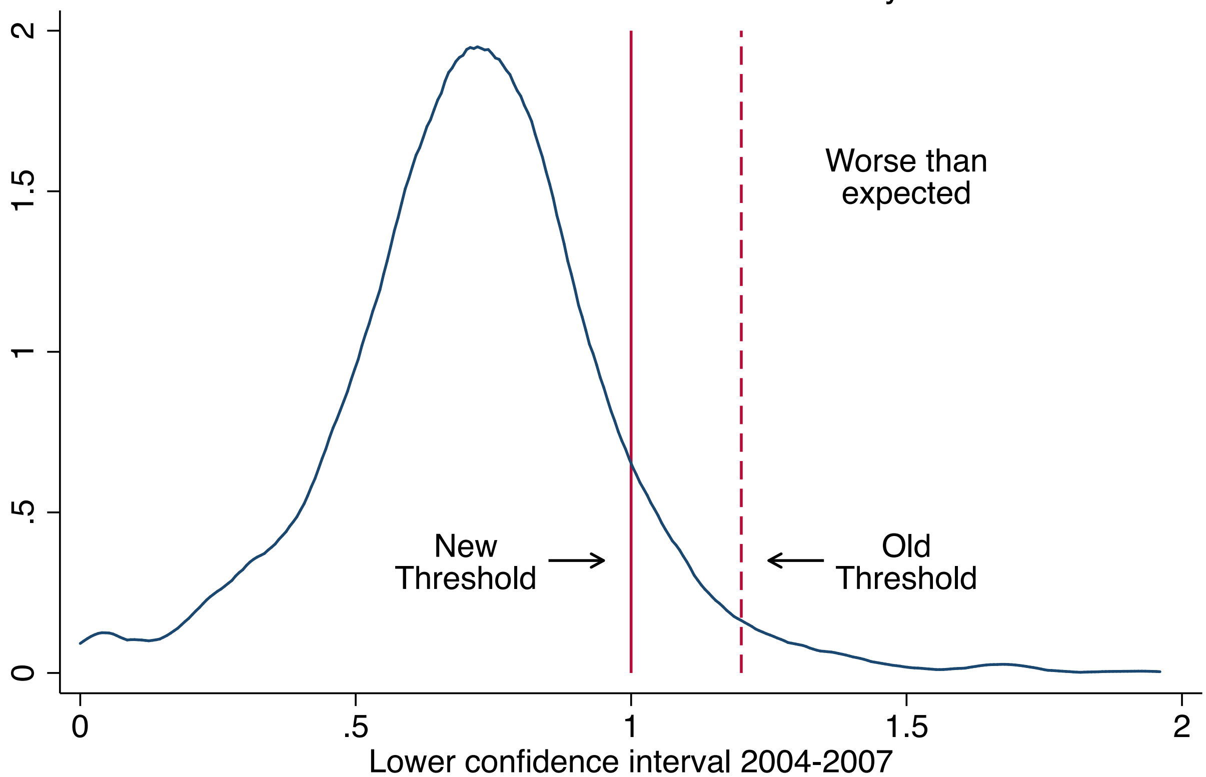


Figure 3: Kernel density of upper confidence interval of the 2004-2007 standardized mortality ratio

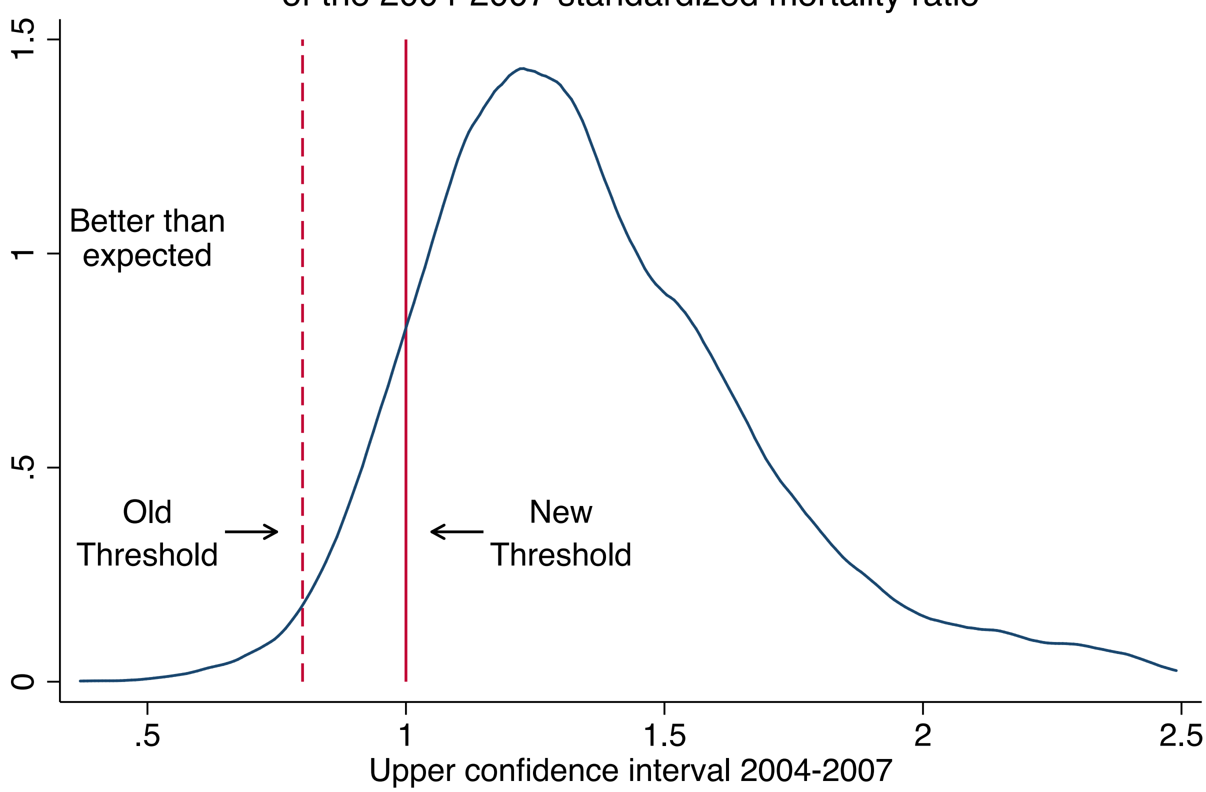
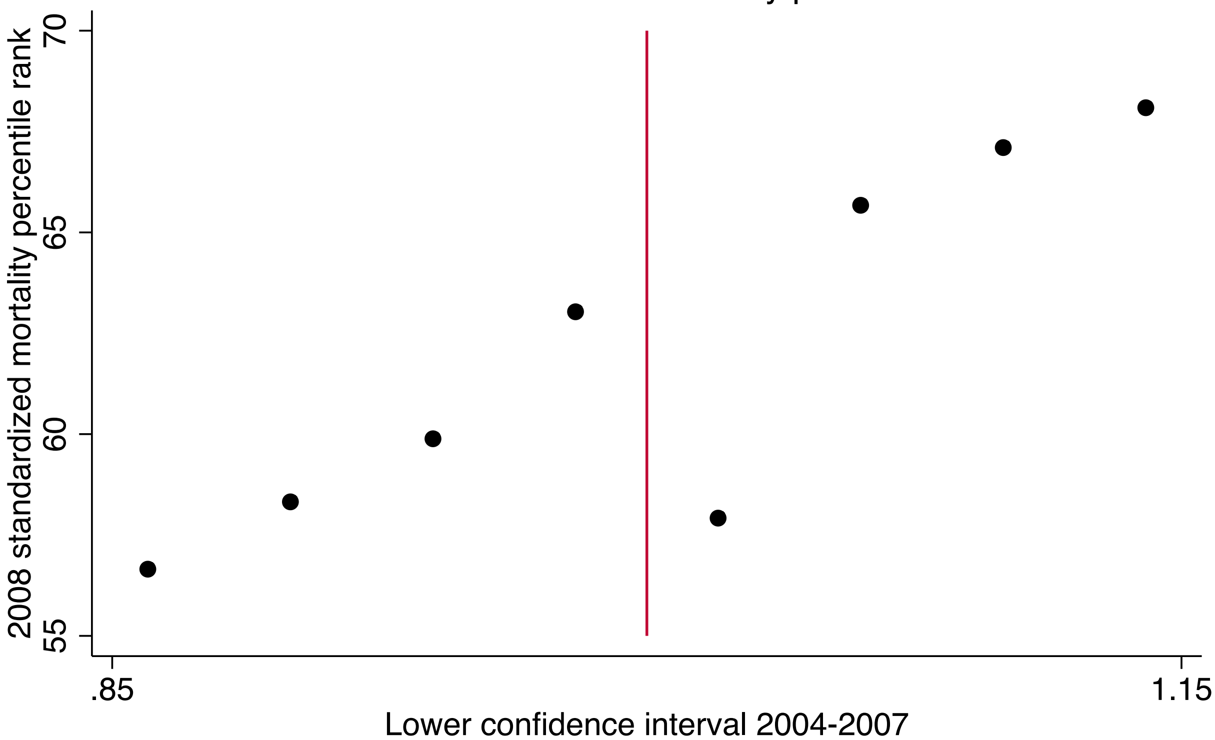
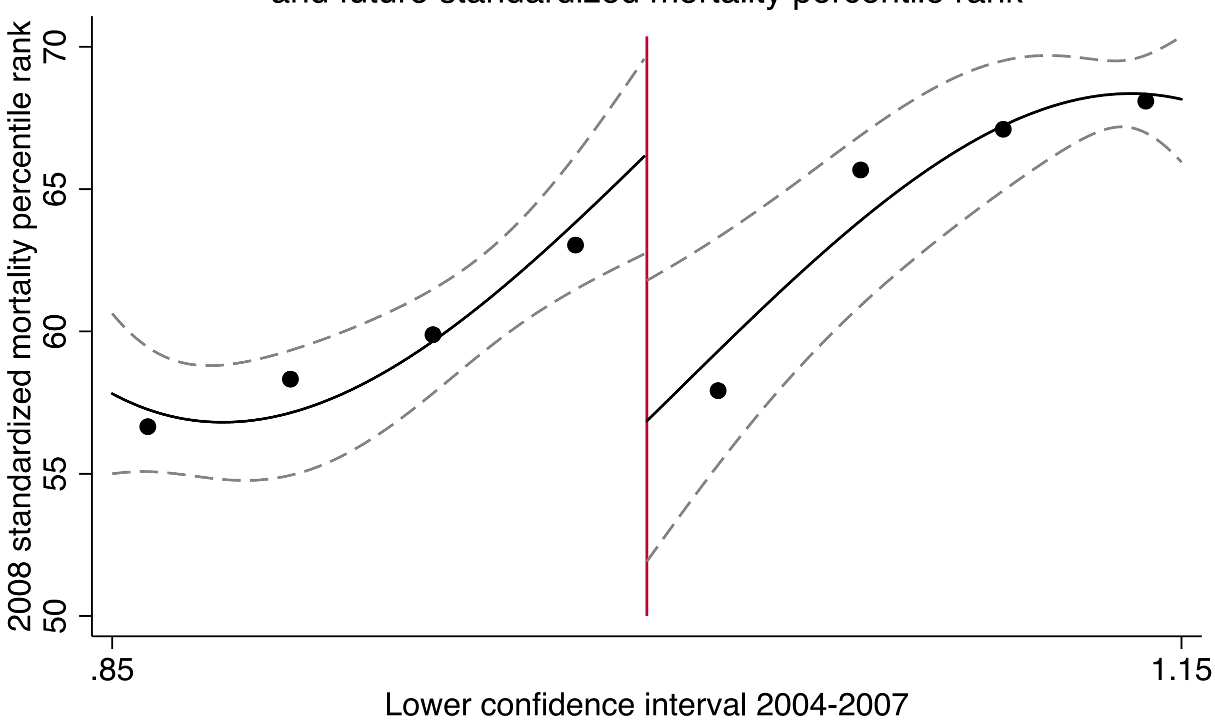


Figure 4: Worse than expected performance threshold
and future standardized mortality percentile rank



Note: Lower rank indicates improved performance

Figure 5: Worse than expected performance threshold
and future standardized mortality percentile rank



Note: Lower rank indicates improved performance. Regression follows specification (1) in table 4 performed on the 8 data points in the graph.

Table 1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Standardized mortality percentile rank 2008	50.033	28.102	1	99	4485
Standardized mortality ratio 2004-2007	1.006	0.302	0	5.14	4665
Worse than expected performance	0.098	0.297	0	1	4665
Better than expected performance	0.103	0.304	0	1	4665
Total stations 2007	17.839	8.339	0	80	4643
For profit status 2007	0.802	0.398	0	1	4665
Chain facility 2007	0.794	0.404	0	1	4665
Number of new patients 2007	21.965	15.829	0	233	4512
Number of new patients 2009	21.101	14.904	0	178	4512
No nephrologist in prior year 2007	0.308	0.206	0	1	4468
No nephrologist in prior year 2009	0.3	0.206	0	1	4464
Left market 2009	0.035	0.183	0	1	4665

Table 2: Observables are balanced at the worse than expected performance threshold

VARIABLES	(1) standardized mortality ratio 2004-2007	(2) Total stations 2007	(3) For profit status 2007	(4) Chain ownership 2007	(5) Above median patient anemia management 2007	(6) Number of new patients 2007	(7) No nephrologist in prior year 2007
Worse than expected performance	-0.010	-0.407	-0.023	-0.064	-0.094	0.502	0.024
	(0.016)	(1.559)	(0.055)	(0.072)	(0.105)	(4.172)	(0.034)
Lower CI polynomial degree	3	3	3	3	3	3	3
Sample size	606	605	606	606	586	590	589

Note: Stars denote significance levels: 99 percent confidence level (***),95 percent confidence level (**), and 90 percent confidence level (*)

Standard errors clustered at the state level. The Sample window range contains observations lower confidence intervals between .9 and .1.

Table 3: Observables are balanced at the better than expected performance threshold

VARIABLES	(1) standardized mortality ratio 2004-2007	(2) Total stations 2007	(3) For profit status 2007	(4) Chain ownership 2007	(5) Above median patient anemia management 2007	(6) Number of new patients 2007	(7) No nephrologist in prior year 2007
Better than expected performance	-0.012	-2.315	0.095	0.138*	-0.128	-1.462	0.031
	(0.016)	(1.697)	(0.083)	(0.081)	(0.084)	(2.664)	(0.039)
Upper CI polynomial degree	3	3	3	3	3	3	3
Sample size	712	710	712	712	688	695	682

Note: Stars denote significance levels: 99 percent confidence level (***),95 percent confidence level (**), and 90 percent confidence level (*)

Standard errors clustered at the state level. The Sample window range contains observations upper confidence intervals between .9 and .1.

Table 4: Threshold status and future standardized mortality percentile ranking

VARIABLES	(1) standardized mortality percentile rank 2008	(2) standardized mortality percentile rank 2008	(3) standardized mortality percentile rank 2008	(4) standardized mortality percentile rank 2008	(5) standardized mortality percentile rank 2008	(6) standardized mortality percentile rank 2008	(7) standardized mortality percentile rank 2008	(8) standardized mortality percentile rank 2008
Worse than expected performance	-14.243***	-15.479***	-9.951**	-9.692**				
	(4.587)	(4.682)	(4.462)	(4.381)				
Better than expected performance					5.762	5.336	1.718	1.961
					(4.112)	(4.225)	(2.753)	(2.797)
Lower CI polynomial degree	3	10	10	10	-	-	-	-
Upper CI polynomial degree	-	-	-	-	3	10	10	10
Control variables	No	No	No	Yes	No	No	No	Yes
Sample window range	.2	.2	.5	.5	.2	.2	.5	.5
Sample size	590	590	1823	1818	692	692	1765	1761

Note: Stars denote significance levels: 99 percent confidence level (***),95 percent confidence level (**), and 90 percent confidence level (*)

Standard errors clustered at the state level.

Table 5: Threshold status and anemia management

VARIABLES	(1) Above median in patient anemia management 2008	(2) Top 10 percentile in patient anemia management 2008	(3) Top 10 percentile in patient anemia management 2008	(4) Above median in patient anemia management 2008	(5) Top 10 percentile in patient anemia management 2008	(6) Top 10 percentile in patient anemia management 2008
Worse than expected performance	-0.022	0.160**	0.084*			
	(0.073)	(0.064)	(0.044)			
Better than expected performance				-0.039	-0.027	0.039
				(0.070)	(0.061)	(0.042)
Lower CI polynomial degree	10	3	10	-	-	-
Upper CI polynomial degree	-	-	-	10	3	10
Sample window range	.5	.2	.5	.5	.2	.5
Sample size	1818	589	1818	1757	688	1757

Note: Stars denote significance levels: 99 percent confidence level (***),95 percent confidence level (**), and 90 percent confidence level (*)

Standard errors clustered at the state level. Anemia is well managed if a patient has a Hemoglobin between 10 and 12 g/dL.

Table 6: Threshold status and the number and characteristics of new patients

VARIABLES	(1) Number of new patients 2009	(2) No nephrologist prior to dialysis 2009	(3) No nephrologist prior to dialysis 2009	(4) Number of new patients 2009	(5) No nephrologist prior to dialysis 2009	(6) No nephrologist prior to dialysis 2009
Worse than expected performance	0.156	0.104**	0.087***			
	(2.816)	(0.044)	(0.031)			
Better than expected performance				1.266	0.032	0.010
				(1.673)	(0.041)	(0.029)
Lower CI polynomial degree	10	3	10	-	-	-
Upper CI polynomial degree	-	-	-	10	3	10
Sample window range	.5	.2	.5	.5	.2	.5
Sample size	1826	586	1819	1770	687	1752

Note: Stars denote significance levels: 99 percent confidence level (***),95 percent confidence level (**), and 90 percent confidence level (*)

Standard errors clustered at the state level.