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

March 2012

Preliminary: Please do not circulate.

Abstract:

We study the impact of information disclosure policies on firm performance by exploiting a policy change that quasi-randomly assigns reputations to firms based on their allocation to coarse performance categories. Dialysis firms are graded on performance by Medicare 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 a mix of improved patient care and strategic patient selection. 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.

We are grateful for helpful comments by Simon Board, Vanessa Burbano, Christopher Carpenter, Phillip Leslie, Steve Lippman, Moritz Meyer-Ter-Vehn, Lamar Pierce and seminar participants at University of Texas Austin, UCLA, UC Irvine and UC Berkeley. Nicholas Ross provided outstanding research assistance.

I. Introduction

Over the past fifty years, extensive literature in economics has highlighted the central role played by information when it comes to decision-making.[1] A central insight gleaned from much of this work is that improving the quality and transparency of information leads to more desirable social outcomes.[2] This finding has spurred the development of quality disclosure programs across various settings aimed at increasing the quality and quantity of information available to consumers. By and large, the academic literature examining the impact of these disclosure programs has found evidence consistent with consumer sorting on quality (Jin and Sorenson 2006; Dafny and Dranove 2008; Hastings and Weinstein 2008) as well as improvements in product quality (Jin and Leslie 2003).[3]

The disclosure programs examined by these studies have a common feature: a quality score, or an ordinal ranking, that is uniformly applied to all of the firms or products in a given marketplace.[4] In practice, however, many disclosure programs are designed quite differently. Two distinct types of programs are frequently observed: those that celebrate the exceptional and those that shame the incompetent. As illustrative examples, consider the quality disclosure programs of the Department of Education and the Environmental Working group. The Blue Ribbon Schools Program administered by the Department of Education "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 approximately 6,000 schools have received a blue ribbon designation out of well over 100,000 eligible schools making this a program that recognizes truly outstanding performers. In contrast to such a designation, the Environmental Working Group recently

[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] The ranking system can be based on absolute or relative performance evaluation. For example, restaurant hygiene is graded on an A to F scale and hospitals (and universities) are ranked from 1 to 100.
[5] See the department of education website: <http://www2.ed.gov/programs/nclbbrs/index.html>

compiled a list of their "Ten Most Sugary Cereals",[6] a list of ten products having the highest sugar content amongst all cereals, the so-called "cookie breakfast".

In spite of the prevalence of such disclosure programs focusing on the extremes of the quality distribution, our understanding of the efficacy of such program designs is still quite limited. Are programs designed to expose the extremes of the distribution more effective at motivating seller investments in product quality when compared to disclosure programs that provide a universal ranking? Should design efforts be focused on exposing the worst performers or on praising the best? Consider a university planning to implement a disclosure program designed with an objective of improving teaching quality among the faculty. Identifying the best teachers might provide an incentive for the average professor to emulate these top performers, thus driving up overall classroom ratings. On the other hand, to the extent that shame is a more powerful motivator, disclosing the identities of the professors with the worst classroom performance could prove more effective at improving teaching quality.[7]

In this paper, we investigate the effectiveness of a disclosure program that focuses on identifying the best and worst performing firms in an industry. In particular, we examine a disclosure program implemented by Medicare wherein dialysis facility centers are graded by being classified into one of three coarse performance categories (*as expected, worse than expected,* and *better than expected*) on the basis of patient survival rates.[8] This setting is well suited to our analysis for two reasons. First, it provides us with a clear, unambiguous measure of performance for each facility (risk-adjusted patient mortality rate), as well as several organizational covariates (e.g. staffing mix, ownership, chain membership) that explain variation in performance. Second, an exogenous change in the methodology employed by Medicare to calculate performance categories coupled with sharp category cutoffs form the basis of our identification strategy.

---

[6] From the Environmental Working Group's website:
<http://www.ewg.org/report/sugar_in_childrens_cereals/best_and_worst_cereals>

[7] Of course, one could design a program that identifies both the worst and best performers, and indeed the setting that we examine in this study is one such example. However, to the extent that measuring and reporting performance is costly, resource-constrained organizations might focus on one or the other.

[8] Henceforth in this paper, we use the terms performance and quality interchangeably. We note that quality in this context is measured using patient survival rates.

Dialysis is the most prevalent mode of treatment for patients suffering from End Stage Renal Disease (ESRD), with the majority of patients undergoing treatment at outpatient facilities. Patients on dialysis have high mortality rates with approximately 20 percent of patients dying within the first year of treatment. Additionally, there is considerable dispersion in patient survival rates across facilities underscoring the important role played by quality disclosure in this context. [9] Since 2001, the Center for Medicare and Medicaid Services (CMS) has made available quality information on dialysis facilities through the Dialysis Facility Compare (DFC) tool available on its website. This quality information is presented in the form of three coarse categories based on patient survival rates at the facility: *as expected*, *better than expected*, and *worse than expected*.[10] Because of the coarseness of the categorical ratings, facilities with very similar patient survival rates could receive very different grades. For example, a facility that had a 90th percentile national ranking of patient mortality over the past four years could receive an *as expected* performance grade while a facility at the 91st percentile could be assigned a *worse than expected* grade, despite these facilities being essentially of the same quality. [11]

Identifying the causal impact of quality disclosure on consumer choice and product quality improvements is challenging because one needs to disentangle the impact of disclosing quality information (on, say, market share) from the impact of underlying product quality (which also drives market share). Our identification strategy relies in part on the fact that in 2008, CMS updated the statistical method used to classify facilities into performance categories. This revision, carried out with an objective of better delineating the best and worst performers from the rest, resulted in far fewer facilities being classified as performing *as expected* in 2009, compared to previous years (80 percent vs. 96 percent, respectively). Correspondingly, the number of facilities being rated as

[9] http://www.theatlantic.com/magazine/archive/2010/12/-8220-god-help-you-you-39-re-on-dialysis-8221/8308/. Section III discusses some of the factors underlying this variation in mortality across centers.
[10] It was not until December of 2010 that the underlying quality scores were released to the broader market. We study a period in time where we can see the underlying quality scores but *only the coarse categories were available to all players in the market.*
[11] 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 III.

superior or inferior performers increased sharply. We exploit this policy change, in conjunction with the coarseness of the quality ratings in a regression discontinuity framework to identify the causal impact of information disclosure on (subsequent) facility quality, and consumer choice.

Our findings suggest that a facility receiving a *worse than expected* performance grade for the years 2004-2007 experiences a substantial improvement in future performance in comparison to a facility that is graded *as expected* based on the 2004-2007 data. In contrast, we find that distinguishing facilities between those that performed *better than expected* and *as expected* has little impact on relevant outcomes. Further analysis reveals that the performance improvement is driven to some extent by both process improvement and strategic patient selection. We find that facilities for which 2004 was the worst year in terms of performance (as compared to 2005, 2006 or 2007), experience smaller improvements due to the reduced incentives to improve (given that the 2004 data is dropped from the next year's performance computation).

We document significant demand-side responses to disclosure as well; we show that knowledgeable patients sort away from facilities that are ranked *worse than expected.* Patients that never saw a kidney specialist prior to beginning dialysis were strongly more likely to join a *worse than expected* facility. Overall, the individual consumer seems to respond strongly to negative information irrespective of facility quality.

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.

II. Information Disclosure and Its Impact on the Dialysis Industry

This study builds on an extensive body of research that examines the impact of quality disclosure programs (e.g. rankings or report cards) across a variety of settings, within and outside of healthcare. We focus on two of the primary findings in the empirical literature that pertain closely to our work. First, at the consumer level, there is substantial evidence that quality disclosure leads to vertical sorting. In other words, consumers are more likely to choose higher quality products post disclosure, when all else is equal. Hastings and Weinstein (2008) provide evidence of such sorting among parents choosing public schools for their children. Jin and Sorenson (2006), Dafny and Dranove (2008), Scanlon *et al.* (2002) show similar findings among enrollees choosing health plans. Consumer response to information disclosure has also been shown to be moderated by a variety of factors. Wedig and Tai-Seale (2002) find a strong response to quality disclosure among first-time enrollees in health plans, while Dranove and Sfekas (2008) show that consumers responded to hospital rankings only when these rankings differed significantly from prior beliefs.

A number of recent studies have also pointed to the role played by consumer attention in determining the overall response to quality information. To the extent that consumers do not incorporate all available information into their decision-making process, the salience of the information presented becomes an important factor in driving consumption choices. For example, DellaVigna and Pollet (2006) find that earnings announcements made on Fridays lead to much weaker stock price responses due to consumer inattention, while Pope (2006) finds that consumers respond to changes in rankings of hospitals in the U.S. News and World Reports rankings, even after controlling for the underlying quality. These findings imply that choosing an appropriate design for quality disclosure programs is central to achieving desirable social outcomes.

The second major finding documented by various studies in this literature is that disclosure leads to sellers investing in 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. Analogous to the variation in consumer response to quality, seller response to quality has also been shown to be heterogeneous with respect to dimensions like competitiveness of the local market. In a study about the impact of disclosure on nursing homes, Chen (2008) finds that nursing homes in more competitive markets experienced larger improvements in quality post disclosure.

This study makes two important contributions to this literature. First, we make use of a regression discontinuity design that allows for sharp identification of the causal impact of information disclosure.[12] Second, the setting of the study enables us to examine the effect of disclosure programs that identify firms on either end of the quality distribution, and our results thus shed light on the optimal design of disclosure programs. As discussed earlier, disclosure programs that classify participants into broad performance buckets are quite common in practice.

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. Conversely, 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. For example, Mas (2006) finds that arrest rates fall when police officers receive pay raises that fall short of their expectations, but the corresponding effect is much smaller when they receive raises that exceed expectations. This suggests that singling out the worst performers can provide a meaningful way to encourage improvements in product quality.

III. Institutional Setting and Data

A. End Stage Renal Disease

[12] Luca(2011) employs a similar methodology to estimate the impact of consumer reviews on restaurant revenue.

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 organ transplantation.[13] Most physicians view transplantation as the preferred mode of 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 kidneys available for transplantation is heavily outnumbered by the number of patients suffering from ESRD leading to a major organ 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.

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

[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/
[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

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. It helps ESRD patients live longer but is not intended as a permanent cure for kidney failure. There are two major categories of dialysis, based on the approach undertaken to remove 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 approximately 5,000 dialysis centers three times a week, with each treatment lasting three to five hours. Over 90 percent of these centers are freestanding facilities and in addition to dialysis services, may provide lab testing and drug infusion services. A typical center provides approximately 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

[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)

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 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.

[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.

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. At the time of introduction of the DFC, based on recommendations from a Consumer Information Workgroup, CMS decided not to report actual patient survival rates (as measured by facility SMRs) but instead reported which of the three coarse categories the facility belonged to, based on where its SMR fell in the nationwide distribution. These categories were denoted *as expected*, *better than expected*, and *worse than expected*. [23] The data used to assign facilities into these categories comes from the National Kidney Foundation which uses a four-year estimation window in order to calculate facility SMRs. For example, performance categories reported by the DFC in 2008 are based off patient survival rates (SMRs) at the facility for the years 2004-2007. CMS uploads the new performance categories to the DFC website in November of each year. [24]

In December 2010, Propublica.org, an independent, non-profit investigative news outlet, made available to the public on their website the precise underlying mortality data for all four-year windows between 2002 and 2009.[25] 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 for patients to adequately compare facilities on the basis of quality.[26]

---

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

[24] In section IV (B), we provide more information on the exact timeline of events relevant to the analysis.

[25] 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.

[26]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*

Using the data uploaded on Propublica.org, we constructed a dataset containing facility level SMRs (raw scores, as well as the coarse categories reported by CMS) as well as a number of facility characteristics for the years 2002-2009 and use the time period 2008-2009 corresponding to the CMS policy change for our analysis.[27] Given that mortality rates are computed using four-year windows, the sample 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. For each of these firms, we obtained information on their past and future patient survival rates, organizational form, ownership, 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.[28] 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 based on interviews with dialysis administrators, Powe *et al.* (2002) report that patient education, staffing ratios, and wage levels 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

---

that Innovative Renal's average annual death rate—after factoring in patient demographics and 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."

[27] These data are available on the Propublica website in the form of reports (in Adobe Acrobat pdf format) for each facility for each year. In order to construct a dataset, we downloaded the entire set of reports from the website (there are ~4500 facilities in the US and the data span nine years from 2002-2010) and then read the relevant variables (e.g. annual facility performance, facility characteristics, patient characteristics, patient volumes) into Microsoft Excel, and subsequently, Stata.

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

associated with more engaged patients, and better communication and coordination between physicians and staff.

Table 1, Table 2 and Table 3 present summary statistics for the 4500+ dialysis facilities in our sample in each of 2007, 2008 and 2009. A majority of these facilities (80 percent) are associated with for-profit ownership and are affiliated with a national or regional chain. Each facility treats over 110 patients on average, with about a fifth of these patients being new to the facility each year. Nearly 70 percent of these patients are referred to the 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.

IV. Estimating the Causal Impact of Information Disclosure on Performance

Establishing the causal impact of quality disclosure poses a difficult challenge. In the absence of a natural experiment (e.g. the policy change exploited by Jin and Leslie 2003), the common methodology in the existing literature has 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.[29] This approach is limited by the fact 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

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

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 driven by 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 differences-in-differences designs can substantially overstate the impact of disclosure programs.[30]

Our identification strategy relies on an exogenous change in methodology employed by CMS in 2008 that led to a dramatic change in the distribution of facilities across performance categories. We exploit this change in conjunction with a regression discontinuity design that exploits the coarseness in performance categories. The following subsections describe the main parts of our empirical approach in greater detail: the change in methodology employed by the CMS to assign performance categories, the timeline underlying the analysis and the intuition behind the regression discontinuity design.

A. *The Change in CMS Methodology*

When the DFC 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 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 was 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.

[30] 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.

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".[31] 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*.[32]

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&3** uses data from the 2004-2007 timeframe to illustrate the impact of the change in thresholds on the classification of dialysis facilities into performance categories. The unexpected "shock" received by many facilities to their reputations provides an ideal setting to examine how sellers respond to quality disclosure. In addition, focusing on the impact of the performance grades released in 2008 (that were assigned based on the 2004-2007 four-year window) ensures that we have sufficient variation in the distribution of firms across performance categories to effectively use the proposed regression discontinuity design.

B. The Timeline of Information Disclosure

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 as follows:

[31] 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.

[32] Facilities with three or fewer deaths are not included in the classification.

- *June 2008*: Each facility receives a Dialysis Facility Report (DFR) from CMS that contains information on 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.[33] Facility level performance data from 2008 would be the first year to show a response to this information.[34]

- *November 2008*: The *Dialysis Facility Compare* website is updated with information on each facility's performance using the 2004-2007 four-year window. Patients looking to undergo dialysis treatments (or currently being treated) learn about the coarse performance category of each facility in their choice set. 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 made 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).

C. The Regression Discontinuity Design

[33] 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 unadjusted patient survival rates alone.
[34] Technically, this effect would be reflected in facility performance only for the latter half of the year since the reports are distributed only in June. However, we only observe SMRs on an annual basis; all our analyses are therefore carried out at the facility-year level.

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."[35] The fundamental identifying assumption is that close to a threshold of interest all other characteristics and choices that could influence an 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 the performance of each facility in 2008 (on the y-axis) and the lower confidence interval (95th percent) of the 2004-2007 SMR on the x-axis. Recall that firms are classified into performance categories based on the x-axis; the vertical line in the graph (drawn at the point where the x-axis equals 1) denotes this classification with facilities on the right side of the line being classified as *worse than expected,* and facilities on the left receiving an *as expected* grade. 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.[36] 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.

The figure provides compelling evidence of a discontinuity precisely at the point where a firm is classified as performing *as expected* and where it is classified as performing *worse than expected*. A firm that just barely falls in the *worse than expected* category is much more likely to have substantially improved performance in 2008 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

[35] 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.

[36] This percentile ranking only uses mortality data from 2008.

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:

$$\text{(1)} \qquad \begin{aligned} Outcome_i = \alpha + \beta * Threshold_i + \gamma * \ polynomial\ of\ CI_i \\ + [\lambda * polynomial\ of\ CI_i * Threshold_i] + [\partial * X_i] + \varepsilon_i \end{aligned}$$

We estimate this equation as a pooled regression on a sample window that extends on both sides of each of the thresholds (i.e., one specification for the threshold between *worse than expected* and *as expected*, and one specification for the threshold between *as expected* and *better than expected*). The primary predictor in these specifications, *Threshold* is an indicator variable which is set to 1 if the facility is classified as performing *better than expected* (or *worse than expected* in the corresponding set of regressions), based on mortality data from 2004-2007.[37] In our analyses examining the impact of disclosure on seller quality, we use the national SMR percentile rank of the facility in 2008 as the dependent variable, *Outcome*. We include polynomial functions of the upper (lower) confidence interval of the SMR in our specifications in order to control for the underlying trend and allow *Threshold* to identify the discontinuous break in the data. As per Lee and Lemieux (2010), we estimate separate regressions on each side of the threshold by allowing for interactions between *Threshold* and the polynomial terms. **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

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

equals 1. In section V.D, we examine robustness of the main results to including polynomials of different orders, as well as to varying the width of the sample window.

We also include control variables in some specifications (denoted by X_i in brackets); these include the facility SMR as computed using the 2004-2007 data (to control for underlying quality), age of the facility (to account for experience-related effects on outcomes), as well as other organizational characteristics measured in 2007 (indicators for chain affiliation and for-profit ownership and the number of total patients, facility size as measured by the number of stations, and competitive intensity of the market as measured by the number of facilities within a 1 mile radius). Inclusion of these control variables should not affect the estimate of interest (if the design is valid) but leads to more precise estimates under certain conditions (Lee and Lemieux, 2010). In addition, these control variables are useful in establishing the validity of the regression discontinuity design, as discussed in the next section.

V. How Do Firms React to Quality Disclosure?

A. Establishing Validity of the Regression Discontinuity Design

In order to establish validity of the Regression Discontinuity design, we demonstrate that *Threshold* is not a significant predictor of the baseline control variables in our model (which are measured in 2007) thus confirming that the assignment of facilities into performance categories is indeed randomized around the cutoffs. We implement this test and report the results in **Figure 6**. Essentially, we estimate separate regressions (analogous to equation 1) in which we use each of the controls as the main dependent variable. The figure lists the t-statistics on the *worse than expected* and the *better than expected* indicators from each of these regressions; we then plot the kernel density of these t-statistics in the Figure and overlay the Normal distribution on top. As can be seen, the distribution of the t-statistics is quite similar to that of the Normal distribution, and does not exhibit any excess dispersion.

B. The Impact of Information Disclosure on Performance

Table 4 presents coefficient estimates of our primary specification examining the impact of being assigned to the *worse (or better) than expected* performance categories on subsequent performance. The estimates in columns 1-4 represent the effect of being assigned to the *worse than expected* category*,* while columns 5-8 present results from specifications that include the *better than expected* indicator. As indicated in equation (1), in all specifications examining the impact of being classified *worse (better) than expected*, we include a 3rd order polynomial of the lower (upper) confidence interval as a control variable. In our base specifications, the estimates are calculated using a sample window of width 0.25 around the threshold.[38]

The coefficient in Column 1 provides strong evidence that a negative report (i.e. a *worse than expected* grade) causes a firm to substantially improve its patient survival rate in 2008. 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 an 11-point percentile improvement in the 2008 SMR percentile ranking. The magnitude of the effect is larger (19-point percentile improvement) when we include interaction terms between *Threshold* and the polynomial terms in column 2. In columns 3 and 4, we show that this result is robust to two major factors that influence regression discontinuity estimates: the size of the sample window and the addition of control variables. 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 789 to 1823. While the coefficient drops in magnitude, the economic significance is still substantial. In column 4, we include the

[38] This corresponds to using observations with lower confidence intervals of the SMR lying between 0.875 and 1.15. For *better than expected* threshold regressions, a sample window range of 0.2 would imply that the upper confidence interval of the SMR would fall between 0.875 and 1.15.

entire vector of control variables referenced above. The coefficient on the indicator is effectively unchanged, suggesting that the treatment is randomly assigned.[39]

In columns 5-8, we present results from similar specifications that examine the 2008 SMR percentile ranking of firms who just barely exceed the *better than expected* threshold. In contrast to the earlier results, the coefficients are much smaller in magnitude and none of them are statistically significant. Furthermore, the coefficient in column 5 is statistically different from the coefficient reported in column 1. Taken together, the results in Table 4 suggest that firms respond to disclosure of 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 displayed by firms receiving "bad" news from information disclosure is economically significant.

What exactly do these improvements in percentile rankings imply in terms of decreases in patient mortality? As an illustrative example from our data, consider the facilities DCI Boston and Satellite Dialysis Round Rock that treated 138 and 130 patients, respectively, in 2008. The total number of recorded patient deaths in facility 1 was 14, while facility 2 experienced 15 patient deaths (both compared to 19.54 expected patient deaths per 100 patient-years, which forms the basis for computation of the SMR in 2008). Based on these 2008 SMR figures, DCI Boston had a percentile rank of 24, while Satellite Dialysis Round Rock had a percentile rank of 45. This example demonstrates the wide fluctuation in short-term performance (measured as rankings from a single year of mortality data) that can result from one or two patient deaths.[40] This likely implies that the quality improvement seen in our specifications arises from facilities implementing changes that result in 1-2 fewer patient deaths. We examine possible mechanisms underlying this improvement in the next subsection.

C. Do Facilities "Screen" Their Way to Better Performance?

[39] While it is surprising that the facility SMR (based on 2004-2007 data) is not a significant predictor of facility performance in 2008, we find (in unreported results) that there is indeed a strong and significant association between these measures once we exclude the *Threshold* indicator from the model.

[40] This is also the reason behind the National Kidney Foundation's decision to use four-year averages while computing performance categories

As documented earlier, the improvement in short-term quality seen at the facilities rated *worse than expected* is economically meaningful. We now carry out analyses that investigate whether this drop in patient mortality is "real" (i.e. follows from investment in process improvement or staffing) or is driven by favorable patient selection. We attempt to disentangle these possible explanations by constructing various proxies for substantive investment in process improvement and for patient selection. In other words, we estimate specifications analogous to equation (1), and use these measures as the dependent variable, $Outcome_i$.

The results from these specifications are presented in **Table 5**; note that each cell in the table corresponds to a regression coefficient. In particular, each cell represents the coefficient on the *worse (better) than expected* indicator when the corresponding row variable is used as the dependent variable, $Outcome_i$. Columns 1-3 present results from specifications using the *worse than expected* indicator, while columns 4-6 present results using the *better than expected* indicator. Across columns, we extend the base specification by including interactions between *Threshold* and the polynomial controls, and by varying the size of the sample window (from 0.25 to 0.5 around the cutoff).

We begin by examining whether the improvement in patient survival rates is driven by process improvements at the facility in terms of higher Urea Reduction Ratios (URR). URR measures the extent to which dialysis is successful in removing urea (a waste product) from the blood, and is used as a proxy for treatment effectiveness. In the data, URR is measured as a percentile ranking; in particular, the URR percentile rank represents the percent of facilities with smaller proportion of patients that have a URR of at least 65%. In other words, a facility with URR percentile rank of 90 would be behind only 10 percent of facilities in terms of the proportion of patients with URR of at least 65%. A facility could theoretically drive up URRs by carrying out longer or more frequent dialysis treatments. The results from specifications using URR percentile rank as the dependent variable however show that URRs were effectively unchanged as a result of quality disclosure.

We next look at whether facilities instituted changes in their staffing ratios in the wake of quality disclosure. Across all specifications, there is no evidence that quality disclosure had any impact on overall staffing ratios; all coefficients are small and imprecisely estimated. This holds true even when we break down the overall staffing ratio and examine the ratio of full-time and part-time dieticians to patients.[41] These results are not entirely surprising given that dialysis providers are divided in their assessment of the importance of optimal staffing ratios for quality improvement (Desai et al 2008).

Given the lack of support in the data for actual quality improvements, we next investigate whether the observed quality improvement can be explained through patient selection. According to a survey conducted by Desai et al. (2009), a majority of dialysis practitioners and staff believe that dialysis facilities engage in "cherry picking" patients by having lower thresholds for turning away patients who were non-compliant with dietary or medication guidelines. Since the performance measure (SMR) is risk-adjusted for a variety of patient co-morbidities, facilities engaging in favorable patient selection to improve their SMR would have to screen patients based on factors not accounted for in the risk-adjustment formula. We focus on one such condition: the average serum albumin level for new patients at a facility in 2008. This measure is strongly correlated with protein-energy malnutrition and has also been found to be a significant predictor of patient mortality (Santos et al. 2003). In addition, to the extent that malnutrition is a visible condition, facilities might find it easier to screen patients on this dimension.

The coefficients (reported in Table 5) on the *worse than expected* indicator in specifications using the new patient average serum albumin level (in 2008) as the dependent variable are imprecisely estimated and economically small in magnitude. Put together with the earlier set of results, we fail to see any evidence of facilities investing in process improvement, or engaging in "cherry picking" of healthier patients *on average*.

[41] Although not reported in the table, results using other staffing components (full-time nurse, part-time nurse, full-time social worker, and part-time social worker) were also noisy and small.

However, to the extent that the incentive to respond to disclosure might vary significantly across facilities, we might observe stronger responses from dialysis facilities that stand to gain the most from doing so.

We exploit the underlying method used to construct the SMR to help us identify facilities that have the strongest incentive to respond to disclosure. In particular, we identify firms for which 2004 was the “worst” year in terms of having the highest annual SMR (among 2004, 2005, 2006 and 2007). Recall that the performance categories for each year are assigned based on computing average SMRs over a time window encompassing the previous four years. This implies that the facilities for which 2004 was the “worst” year would experience a mechanical improvement in performance (in terms of the 2009 categories) relative to other facilities (conditional on 2008 SMR) once the time window shifts to 2005-2008. Based on this reasoning, we characterize such facilities as having a weaker incentive to institute performance improvements in 2008 compared to facilities for which the “worst” year was 2005, 2006 or 2007.

We capture this variation in the incentive to respond to disclosure by including an interaction term between *worse than expected* and an indicator for whether 2004 was the “worst” year for that facility. **Table 6** presents coefficient estimates from specifications analogous to the ones reported in Table 5, but with the inclusion of this interaction term. We only present results for the specifications that include the *worse than expected* indicator, given that we found performance improvements only for firms assigned to that category.

We focus our discussion on outcome variables for which we obtain statistically significant estimates that are robust across our various specifications (varying the size of the sample window and including interactions with the polynomial terms). First, we find significant improvement in URR levels for facilities with the strongest incentive to improve performance – the interaction term is negative, implying that facilities for which 2004 was the “worst” year (i.e. facilities for which incentives were weakest) had much lower URR percentile ranks in 2008. Second, we also find that such facilities increased the number of part-time dieticians on staff, which might be one of the driving factors behind the improvement in URR levels. Taken together, these results provide some support for facilities improving overall performance through investments in staffing and process improvement.

Interestingly, we also find results consistent with incentivized facilities engaging in more extensive patient screening. In particular, the average serum albumin levels for new patients in 2008 were significantly higher (by 0.12 units compared to a mean of 3.12 units) for incentivized facilities; meaning that these facilities were treating patients that were healthier in ways not captured in the risk-adjustment.

In sum, we find that facilities that receive negative information (i.e. a *worse than expected* rating) via quality disclosure display an improvement in short-term performance, as measured by SMR in 2008. This improvement is larger in magnitude among facilities that have a stronger incentive to bring SMRs down in 2008. Finally, our results indicate that the performance improvement was driven by both investments in process improvement as well as increased "cherry picking" of patients.

D. Robustness Checks

We examine robustness of our results to various alternate polynomial orders as well as sample window sizes and present the results in graphical form. **Figure 9** and **Figure 10** demonstrate the robustness of the coefficient estimates in Table 4 (on the *worse than expected* indicator) to varying the order of the control polynomial for a sample window of width 0.25 and 0.5 around the cutoff, respectively. Our results are also robust to varying the width of the sample window from 0.05 to 0.5 for a 3^{rd} and 5^{th} order polynomial, as shown in **Figure 11** and **Figure 12**.

VI. How Does Quality Disclosure Affect Patient Flow?

Table 7 presents coefficients 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 composition of new patients only in 2009. The coefficients in the first two rows present the impact of disclosure on the number of patients and the number of new patients at the facility in 2009. The estimate is noisy, and the large standard errors make it difficult to even pinpoint the direction of the effect. The same holds true for firms just passing the *better than expected* threshold.

We next consider the impact of disclosure on the composition of patients at a facility. In particular, we focus on the proportion of patients who have never consulted a nephrologist prior to starting dialysis. Nephrologists are kidney specialists who help patients manage the course of ESRD. Patients benefit from visiting a nephrologist because 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. The summary statistics in Table 3 indicate that approximately 30 percent of patients did not visit a nephrologist prior to starting dialysis. When using this measure as our outcome variable, we find strong evidence that when a facility is just barely within the *worse than expected* threshold, an additional 9 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 passing the *better than average* threshold on the number of new patients that are relatively less informed.

VII. Conclusions

The evidence in this paper is consistent with information disclosure programs being most effective when the incompetent are shamed. Further, the performance improvement seen in firms (as a response to disclosure) seems to be driven by a mixture of investments in process improvement and strategic patient selection. 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. In addition, our finding on the composition of new patients (i.e. less knowledgeable patients attending poorly rated facilities) suggest that the information on the website is not being utilized by all of the patients seeking dialysis treatment and points to the need for increasing awareness of the *Dialysis Facility Compare* program and improving its design so it is better understood.

Some caveats to our findings are in order. First, our measure of performance improvement is based on short-term facility responses, so our analyses do not fully capture the entire impact on quality, some of which may only manifest over the long term. Second, the dialysis industry has a few unique features (e.g. lack of price competition, capacity constraints among high quality firms), implying that our findings may not translate equally well across all settings. Notwithstanding these caveats, the results from this study clearly imply that designing quality disclosure programs in ways that make the information presented more salient to firms and consumers could lead to substantial welfare improvements.

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.

Desai, A. et al. 2008. "Identifying Best Practices in Dialysis Care: Results of Cognitive Interviews and a National Survey of Dialysis Providers", *Clinical Journal of the American Society of Nephrology,* 3(4), 1066-1076.

Desai, A. et al. 2009. "Is there "Cherry Picking" in the ESRD Program? Perceptions from a Dialysis Provider Survey", *Clinical Journal of the American Society of Nephrology*, 4, 772-777.

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.

Santos, N. et al. 2003. "Is Serum Albumin a Marker of Nutritional Status in Hemodialysis Patients without Evidence of Inflammation?", *Artificial Organs*, 27(7), 681-686.

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.

Table 1: Facility summary statistics

Variable	Observations	Mean	Standard deviation
Standardized Mortality Ratio: 2004-2007	4665	1.01	0.30
Worse than expected performance	4665	0.10	0.30
Better than expected performance	4665	0.10	0.30
Chain affiliated: 2007	4643	0.80	0.40
For-profit: 2007	4643	0.81	0.40
Facility age	4512	12.32	8.61
Total patients: 2007	4512	111.99	67.88
Total stations: 2007	4643	17.84	8.34
In-unit patient : station ratio: 2007	4449	3.76	1.85
Standardized mortality ratio: 2004	4080	1.06	0.43
Worst year for standardized mortality ratio = 2004	4665	0.40	0.49

Table 2: Facility characteristics in 2008

Variable	Observations	Mean	Standard deviation
National standardized mortality percentile rank: 2008	4485	50.03	28.10
New patient average serum albumin: 2008	4309	3.12	0.34
Observed deaths per 100 patient years: 2008	4509	21.53	10.36
Expected deaths per 100 patient year: 2008	4509	21.18	6.18
URR national rank: 2008	4412	54.29	34.37
HG national rank: 2008	4469	49.92	28.49
Average treatment time: 2008	4012	4.36	0.52
Staff : total patients ratio: 2008	4506	0.15	0.12
Full time dietican: total patient ratio: 2008	4409	0.01	0.01
Part time dietican: total patient ratio: 2008	4409	0.01	0.01
Facility drops out: 2010	4665	0.04	0.18
Independent facility 2008 merged with chain in 2010	937	0.07	0.249

Table 3: Patient volume and composition in 2009

Variable	Observations	Mean	Standard deviation
Total patients: 2009	4512	113.78	67.56
% of all patients new to dialysis: 2009	4504	0.19	0.08
% of all patients transferred in: 2009	4504	0.12	0.08
% of all patients transferred out: 2009	4504	0.14	0.09
Total number of new patients: 2009	4512	21.10	14.90
% never seen a nephrologist prior to dialysis: 2009	4464	0.30	0.21
% of patients white: 2009	4464	0.66	0.30
% of patients black: 2009	4464	0.29	0.29
% of patients no insurance: 2009	4464	0.07	0.11

Table 4: The impact of information disclosure on facility performance

	Dependent variable: National standardized mortality percentile rank: 2008							
Independent variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Worse than expected performance	-11.03 (4.40)**	-19.18 (6.29)***	-15.22 (5.30)***	-14.89 (5.22)***				
Better than expected performance					4.40 (3.53)	6.74 (5.30)	2.75 (3.60)	3.07 (3.72)
Standardized Mortality Ratio: 2004-2007				8.01 (7.65)				8.06 (10.07)
Chain affiliated: 2007				-1.27 (1.31)				-1.17 (1.26)
For-profit: 2007				2.84 (2.09)				5.31 (1.37)***
Facility age				-.04 (.09)				.06 (.08)
Total patients: 2007				-.01 (.02)				-.00 (.01)
Total stations: 2007				.20 (.15)				.22 (.12)*
In-unit patient : station ratio: 2007				.94 (.65)				.03 (.70)
Neighboring facilities count within 1 mile				.43 (.52)				.28 (.81)
Order of polynomial for confidence interval	3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd
Threshold X Polynomial for confidence interval		Yes	Yes	Yes		Yes	Yes	Yes
Sample range	.25	.25	.5	.5	.25	.25	.5	.5
Observations	789	789	1823	1808	904	904	1765	1755

Note: *, **, *** indicate significance at the 10%, 5%, and 1% confidence levels, respectively. Parentheses contain standard errors clustered at the state level.

Table 5: Mechanisms underlying performance improvement: Coefficient estimates from various outcome regiressions

	Independent variable: Worse than expected			Independent variable: Better than expected		
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
	Disclosure on patient selection					
New patient average serum albumin: 2008	-.04 (.05)	-.03 (.11)	.03 (.07)	.02 (.07)	.00 (.11)	.04 (.08)
	Disclosure on facility treatment rank					
URR national rank: 2008	-.40 (7.23)	-3.72 (9.84)	1.44 (7.15)	5.87 (5.42)	.77 (8.62)	7.19 (5.78)
	Disclosure on facility staffing					
Staff : total patients ratio: 2008	.005 (.008)	-.003 (.011)	.002 (.014)	.008 (.008)	.014 (.011)	.027 (.018)
Full time dietican: total patient ratio: 2008	-.001 (.001)	-.002 (.002)	-.002 (.001)	.003 (.002)*	.003 (.004)	.004 (.002)
Part time dietican: total patient ratio: 2008	.002 (.001)	.001 (.002)	.003 (.001)*	-.000 (.001)	.003 (.002)	.002 (.002)
Order of polynomial for confidence interval	3rd	3rd	3rd	3rd	3rd	3rd
Threshold X Polynomial for confidence interval		Yes	Yes		Yes	Yes
Sample range	.25	.25	.5	.25	.25	.5

Note: *, **, *** indicate significance at the 10%, 5%, and 1% confidence levels, respectively. Parentheses contain standard errors clustered at the state level.

Table 6: Mechanisms underlying performance improvement: Coefficient estimates from incentivized facilities

	Independent variable: 2004 Worst SMR dummy X Worse than expected mortailty			
Dependent variable	(1)	(2)	(3)	(4)
	No Incentives: Disclosure on performance and patient selection			
National standardized mortality percentile rank: 2008	.11 (4.99)	.62 (6.48)	6.65 (9.42)	5.80 (6.75)
New patient average serum albumin: 2008	-.12 (.05)***	-.12 (.05)**	-.21 (.06)***	-.13 (.05)**
	Disclosure on facility treatment rank			
URR national rank: 2008	-14.31 (4.54)***	-13.69 (4.50)***	-10.49 (7.22)	-11.24 (5.18)**
	Disclosure on facility staffing			
Staff : total patients ratio: 2008	.005 (.007)	.006 (.007)	-.005 (.009)	-.025 (.018)
Full time dietican: total patient ratio: 2008	.000 (.001)	.000 (.001)	.000 (.001)	.001 (.001)
Part time dietican: total patient ratio: 2008	-.002 (.001)*	-.002 (.001)	-.003 (.002)**	-.004 (.002)*
Order of polynomial for confidence interval	3rd	3rd	3rd	3rd
Threshold X Polynomial for confidence interval		Yes		
Order of polynomial for distance between 2004 SMR and next worst year	3rd	3rd	3rd	3rd
2004 Worst SMR dummy X Polynomial for 2004 SMR distance		Yes		
Polynomials interacted			Yes	Yes
Sample range	.25	.25	.25	.5

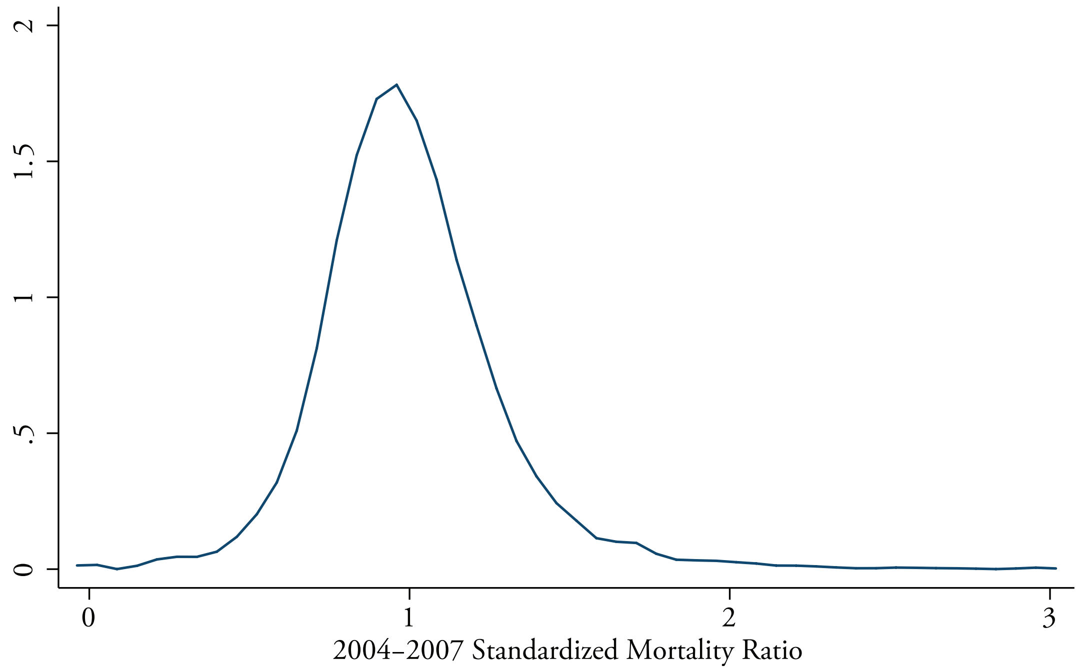
Note: *, **, *** indicate significance at the 10%, 5%, and 1% confidence levels, respectively. Parentheses contain standard errors clustered at the state level.

Table 7: Impact of information disclosure on patient volume and composition

Dependent variable	Independent variable: Worse than expected (1)	(2)	(3)	Independent variable: Better than expected (4)	(5)	(6)
	Disclosure on future patient enrollment and flows					
Total patients: 2009	2.51 (11.44)	17.61 (15.00)	11.38 (13.35)	-7.38 (11.56)	-31.96 (23.98)	-15.99 (13.90)
Total number of new patients: 2009	2.31 (3.34)	8.51 (5.28)	3.19 (3.71)	1.50 (2.26)	-4.98 (4.43)	.00 (2.59)
	Disclosure on future patient pool composition					
% never seen a nephrologist prior to dialysis: 2009	.09 (.03)**	.13 (.05)**	.14 (.04)***	.04 (.03)	.04 (.05)	.04 (.04)
% of patients no insurance: 2009	.04 (.02)*	.02 (.03)	.02 (.02)	.01 (.02)	.02 (.03)	.01 (.03)
Order of polynomial for confidence interval	3rd	3rd	3rd	3rd	3rd	3rd
Threshold X Polynomial for confidence interval		Yes	Yes			
Sample range	.25	.25	.5	.25	.25	.5

Note: *, **, *** indicate significance at the 10%, 5%, and 1% confidence levels, respectively. Parentheses contain standard errors clustered at the state level.

Figure 1: Distribution of dialysis facility performance



Note: Kernel density of the distribution of the 2004-2007 Standardized Mortality Ratio

Figures 2 & 3: Quality reporting thresholds move inward in 2008

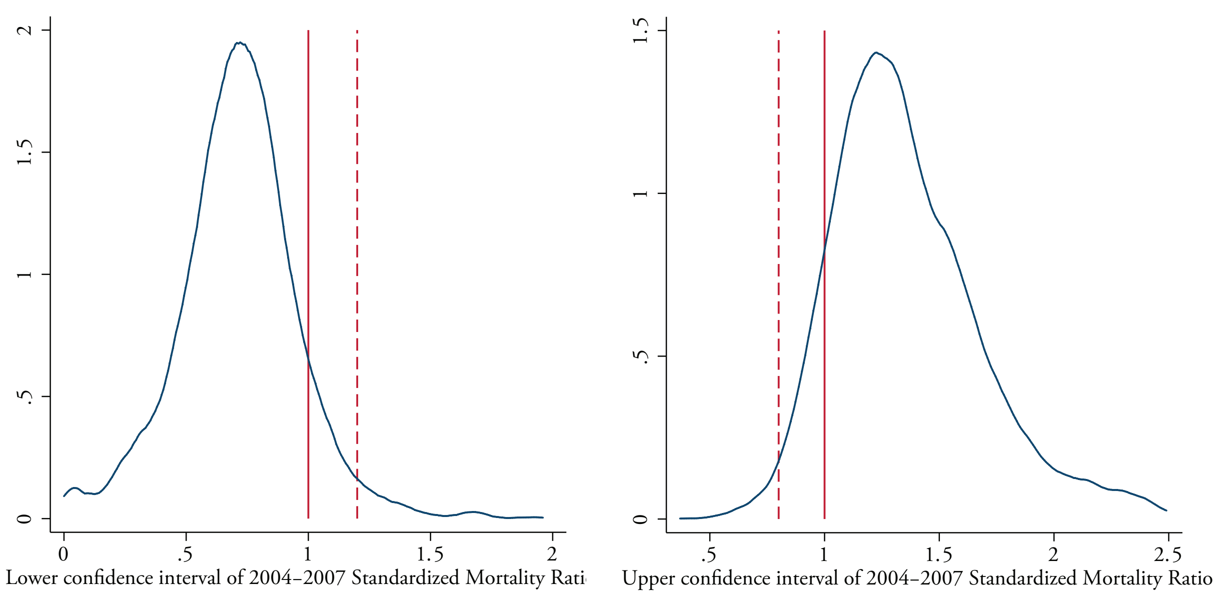


Figure 4: 2008 National standardized mortality percentile rank

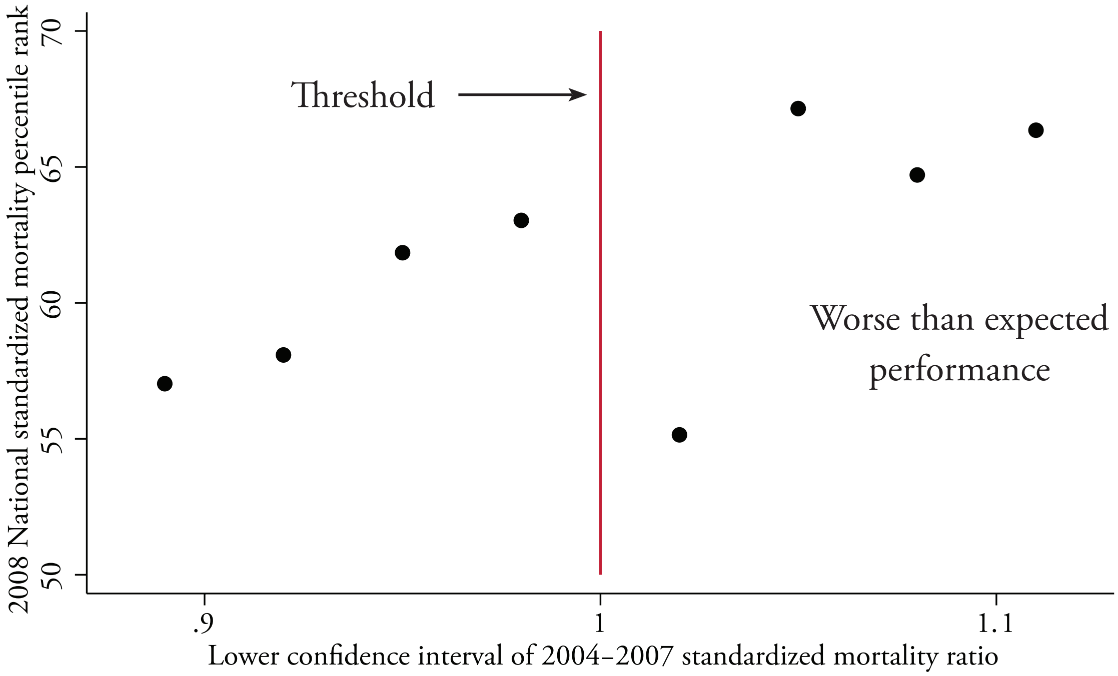
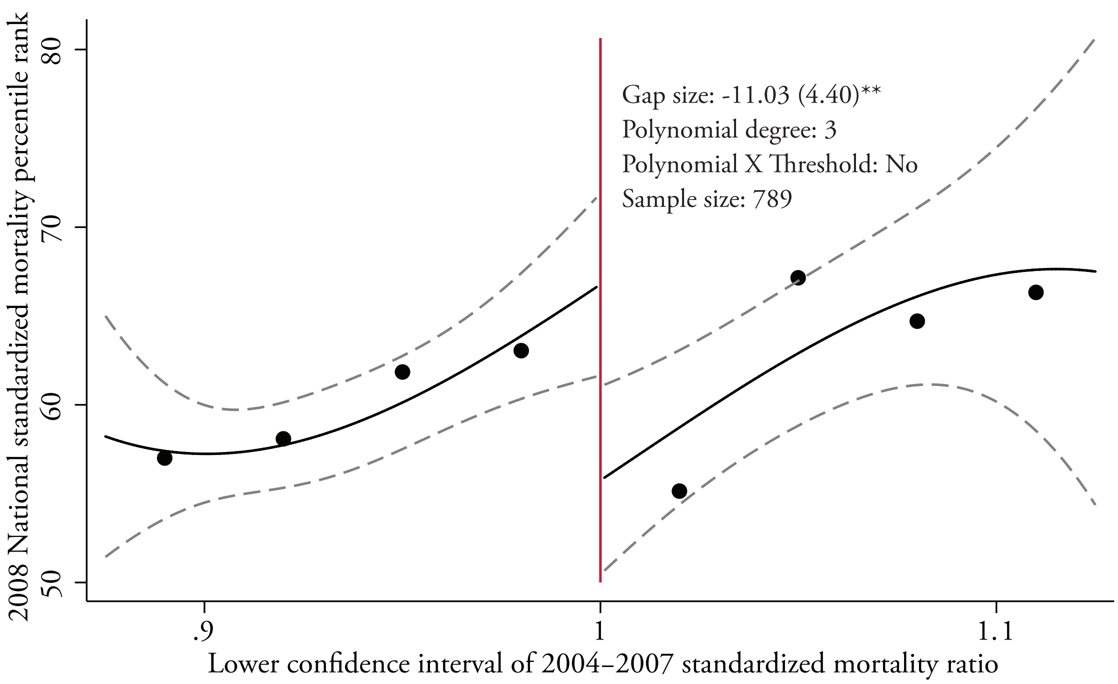
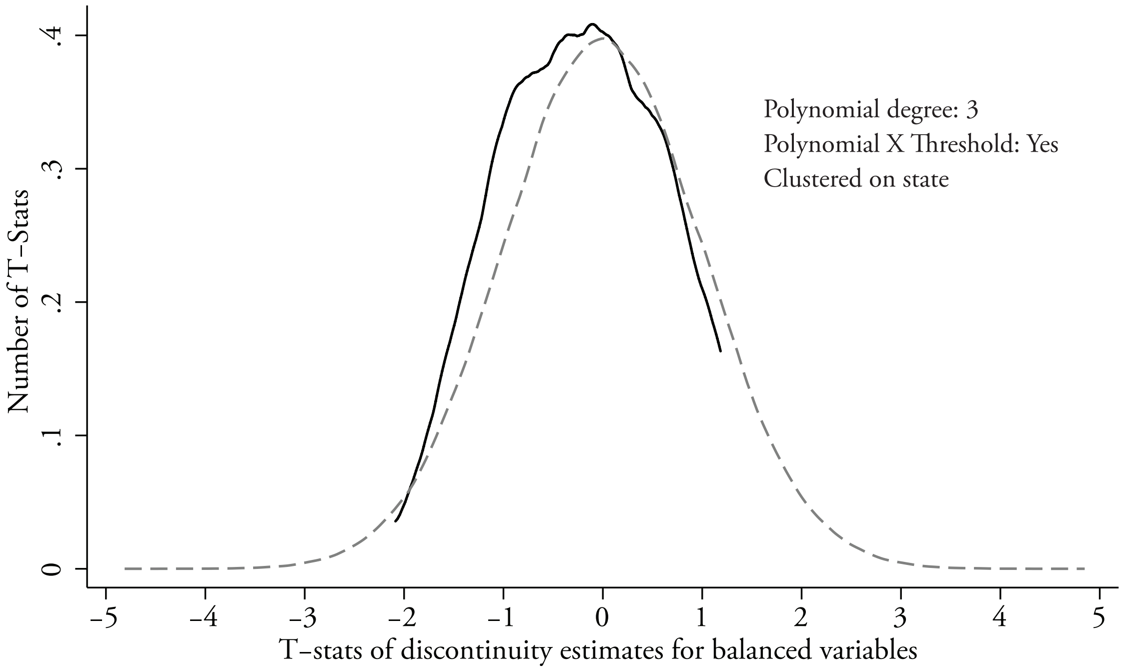


Figure 5: Regression discontinuity estimates for the impact of "worse than expected" rating on subsequent performance



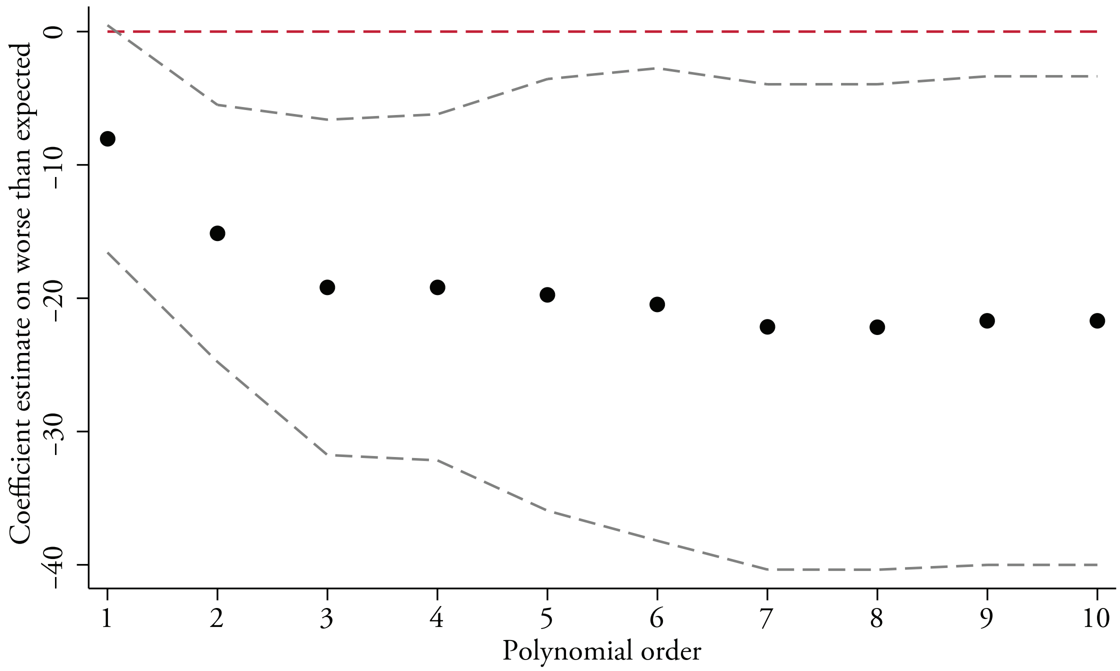
Note: *, **, *** indicate significance at the 10%, 5%, and 1% confidence levels, respectively. Parentheses contain standard errors clustered at the state level.

Figure 6: Kernel density of balancing t-stats do not exhibit excess dispersion



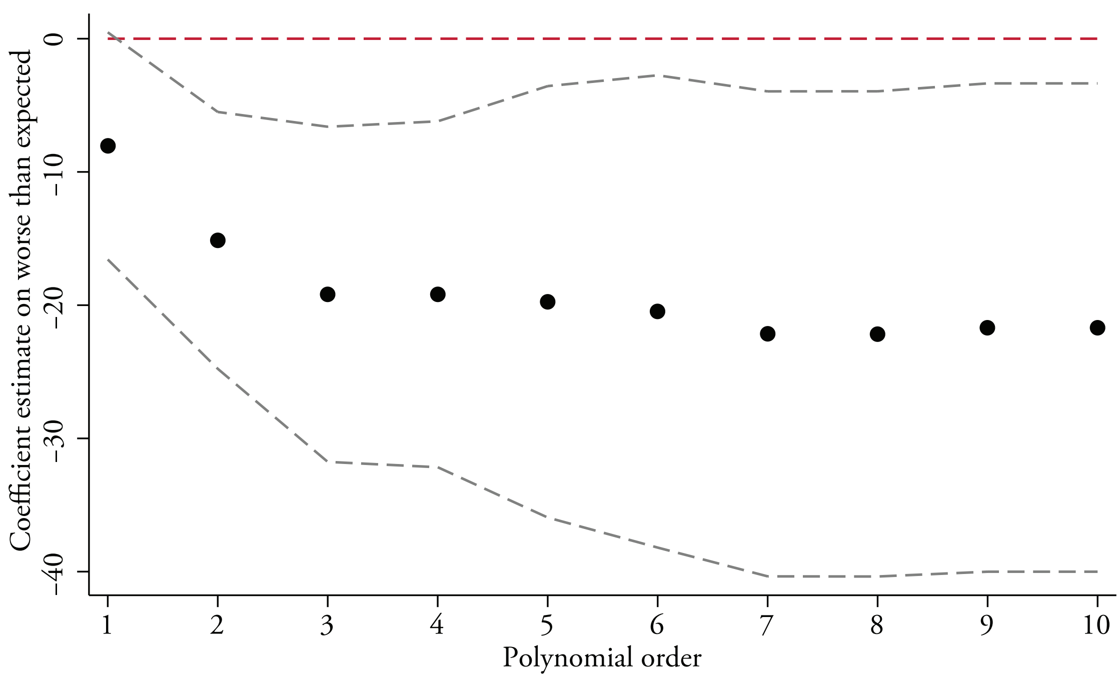
Variable	Worse t-stat	Better t-stat	Variable	Worse t-stat	Better t-stat
Standardized Mortality Ratio: 2004-2007	-.93	-.66	URR national rank: 2007	.49	-1.06
Chain affiliated: 2007	.02	1.13	HG national rank: 2007	-1.31	-.78
For-profit: 2007	.70	.07	Staff : total patients ratio: 2007	.15	.56
Facility age	.35	-1.56	Full time dietican: total patient ratio: 2007	-.72	1.10
Total patients: 2007	-.19	-1.23	Part time dietican: total patient ratio: 2007	.50	.74
Total stations: 2007	.16	-1.41	% of all patients new to dialysis: 2007	-1.16	1.18
In-unit patient : station ratio: 2007	-.26	-.07	% of all patients new to transfer in: 2007	.19	-.95
Neighboring facilities count within 1 mile	-.52	-2.09	% of all patients new to transfer out: 2007	-1.29	.23
2004 Worst SMR dummy	1.03	-.46	Total number of new patients: 2007	-.26	-.49
New patient average serum albumin: 2007	-.96	.00	% never seen a nephrologist prior to dialysis: 2007	1.18	,87
Observed deaths per 100 patient years: 2007	-.74	-.42	% of patients black: 2007	-.45	-1.07
Expected deaths per 100 patient year: 2007	-.47	.52	% of patients no insurance: 2007	.21	.11

Figure 7: Robustness to polynomial order for sample window size = .25



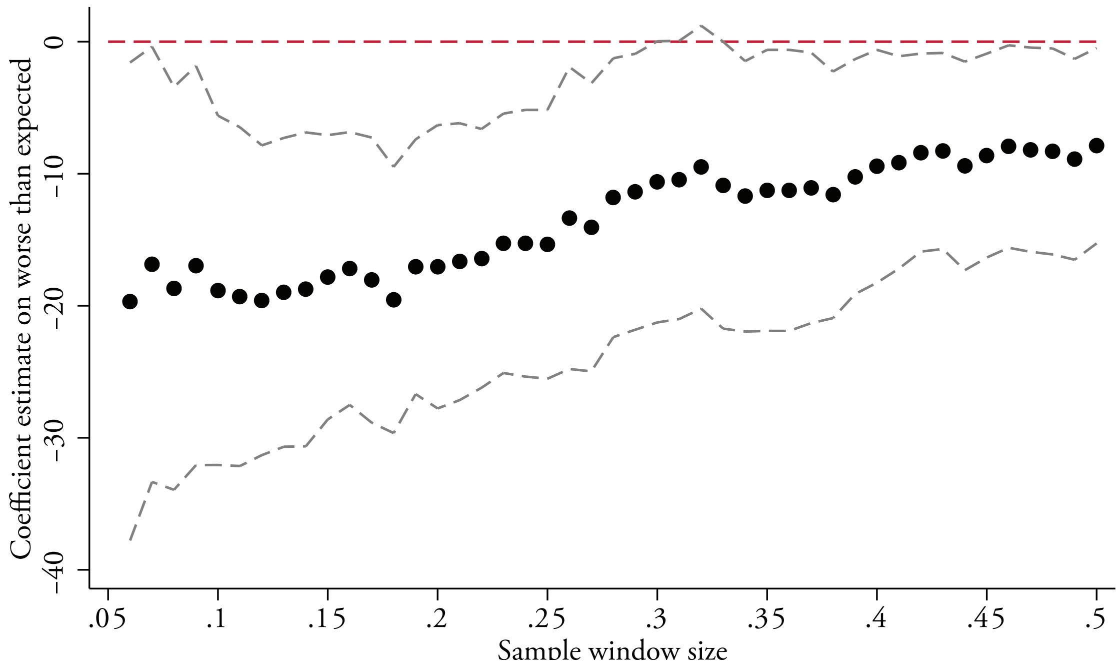
Note: This specification is based on a worse than expected dummy variable interacted with N[th] order polynomial similar to the specifications in table 4 with a .25 window size. Dashed lines indicated 95% confidence interval.

Figure 8: Robustness to polynomial order for sample window size = .5



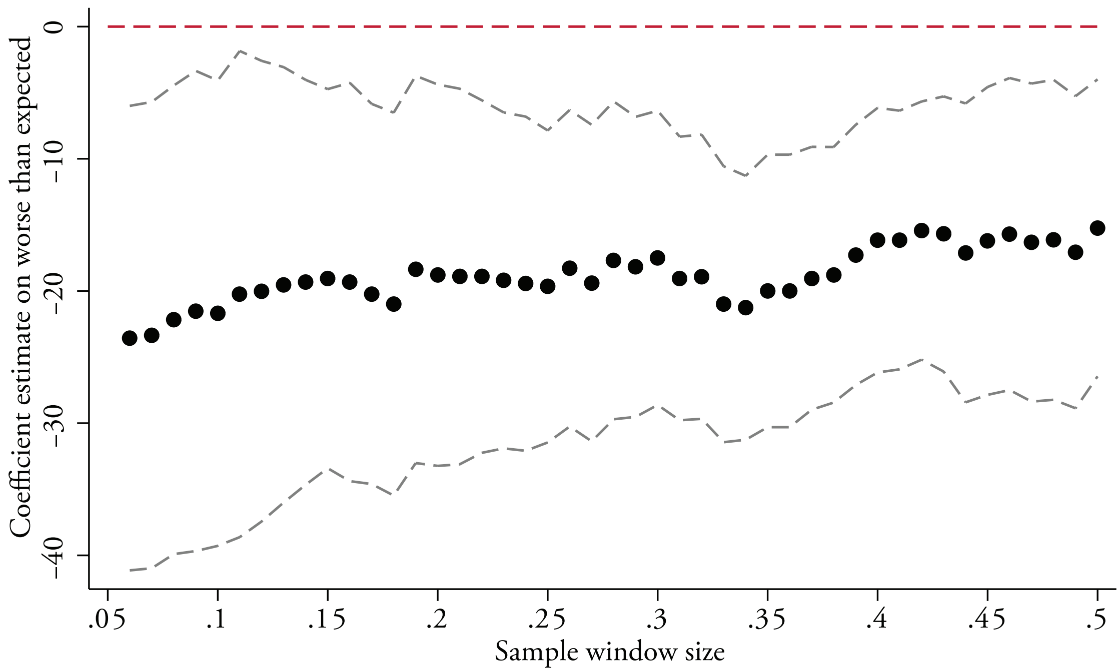
Note: This specification is based on a worse than expected dummy variable interacted with N[th] order polynomial similar to the specifications in table 4 with a .5 window size. Dashed lines indicated 95% confidence interval.

Figure 9: Robustness to sample window size for 3rd order polynomial



Note: This specification is based on a worse than expected dummy variable interacted with 3rd order polynomial similar to the specifications in table 4. Dashed lines indicated 95% confidence interval.

Figure 10: Robustness to sample window size for 5th order polynomial



Note: This specification is based on a worse than expected dummy variable interacted with 5th order polynomial similar to the specifications in table 4. Dashed lines indicated 95% confidence interval.