

The Effect of Performance Standards on Healthcare Provider Behavior: Evidence from Kidney Transplantation

Sarah S. Stith*

Department of Economics
University of New Mexico
ssstith@unm.edu

Richard A. Hirth

Department of Health Management and Policy
School of Public Health
University of Michigan
rhirth@umich.edu

August 20, 2015

Abstract

Performance standards are designed to ensure a basic level of quality, and through public reporting of firm performance, encourage firms to compete on quality thus allowing the market to determine the optimal level of quality. In markets with substantial excess demand, however, demand effects may be insufficient to induce any change in firm behavior and enforcement may be required to ensure high quality. Even with enforcement, quality still may not improve at underperforming firms if gaming the system is less costly than improving quality. We test whether information alone or with regulatory enforcement improves outcomes or elicits gaming behavior in our study of 266 kidney transplant centers between 2001 and 2012. In a context of excess demand induced by price controls, we show that information alone has no impact and enforcement may actually increase market inefficiencies; firms respond to costly quality requirements, not by improving quality, but by reducing supply, which exacerbates the disequilibrium between supply and demand, and by cream-skimming, which reduces access to transplantation among sicker patients.

*Corresponding author

We would like to thank the Robert Wood Johnson Foundation Center for Health Policy at the University of New Mexico and the Ross School of Business at the University of Michigan for the financial support necessary to purchase the data used in this paper as well as Tom Buchmueller, Brady Horn, Alan Leichtman, Lauren Nicholas, seminar attendees at IUPUI and the RWJF Center for Health Policy, and participants at the Fifth Biennial Conference of the American Society of Health Economics for helpful comments. In addition, “this work is supported in part by Health Resources and Services Administration contract 234-2005-370011C. The content is the responsibility of the authors alone and does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.” <http://optn.transplant.hrsa.gov/data/citing.asp>

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/jems.12161](https://doi.org/10.1111/jems.12161).

This article is protected by copyright. All rights reserved.

1. Introduction

The use of publicly reported quality data by regulators theoretically should address market failures associated with a lack of competition and incomplete information, thus increasing market efficiency including improvements in quality. Consumers will make better informed choices leading to increased competition among firms and overall quality should increase by avoiding the low quality equilibrium predicted in Akerlof's (1970) paper on used cars. Public and/or private reporting of product or service quality occurs in many sectors (e.g., automobile safety test data are reported by the National Highway Traffic Safety Administration while quality reviews and reliability data are reported by private sector organizations such as the Consumers Union). However, quality reporting and standards are particularly prevalent in healthcare due to the extent of information asymmetry between producers and consumers and the significant (and often irreversible) consequences of receiving poor quality care. More broadly, quality reporting is a key component of a healthcare policy strategy to overcome market failure by empowering consumers to be agents of change in efforts to improve quality and contain costs. When quality standards are enforced by regulators, in addition to public reporting of quality, the incentives for firms to provide high quality services should be further increased. However, demand side responses might decrease with enforcement if consumers can depend on regulators to ensure quality.

Unfortunately, many such regulated markets (e.g., in healthcare) suffer from additional market failures that make the implications of quality reporting and enforcement less obvious. For example, and as we will test in this paper, if the introduction of public quality reporting occurs in a market with significant excess demand, any consumer response will be insufficient to induce a change in firm behavior. Introducing quality enforcement without any change in the price will likely only exacerbate the disparity between supply and demand as firms choose not to serve their less profitable customers. Such cream-skimming behavior

could be further exacerbated in the presence of insufficient risk adjustment or if the quality measure is difficult for firms to affect.

In transplantation, some form of quality monitoring, based on one year post-transplant patient and organ graft survival, has existed since 1991. Beginning in 2001, this information has been publicly available online through the Scientific Registry of Transplant Recipients (SRTR) with new reports issued every six months for all transplant programs.¹ The standards are monitored by the Department of Health and Human Services through the Organ Procurement and Transplantation Network (OPTN). The OPTN generally has taken an advisory rather than punitive role with transplant centers that failed to meet the survival standards. In response to concerns about a lack of enforcement by the OPTN, the Centers for Medicare and Medicaid Services (CMS) began enforcing the OPTN quality standards with the threat of an expensive audit and possible closure of noncompliant transplant centers, starting on June 28, 2007 (CMS, 2007). Other insurers also require that centers meet these or stricter standards in order to remain part of their provider network.

Using a simple model of the supply and demand for kidney transplants, we predict that information alone will be insufficient to cause a change in supplier behavior, primarily due to the existence of significant excess demand induced by a price ceiling. In the presence of enforcement, we expect centers to improve outcomes primarily through gaming behaviors; they will do so both when noncompliant and preemptively when approaching noncompliance.

As our model predicts, we find no impact from information alone. All our results are only statistically significant after June 28, 2007, when CMS increased its oversight of transplantation. Enforcement leads to more cautious behavior at the average transplant center and induces a significant gaming response in noncompliant centers or centers approaching noncompliance. Our results indicate, however, that despite a strategic response from transplant centers, no evidence of a statistically significant improvement in

¹ www.srtr.org

reported post-transplant survival rates exists. We only see evidence of changes in “gaming” outcomes. Noncompliant centers decrease transplant rates for both deceased and live donor transplants and increase the number of patients removed from the waitlist as too sick for transplant or transferred to another center. We also find evidence of preemptive behavior based on the publicly posted information in the CMS enforcement period; centers respond to almost breaching survival thresholds in a similar but muted manner relative to their response to reports documenting regulatory noncompliance. Our theoretical prediction that any demand shift will not impact the quantity or quality of transplants (due to excess demand) and the lack of any effect on outcomes prior to the introduction of CMS enforcement suggests that transplant center responses are the primary drivers of our results and affecting demand does not have a meaningful impact on providers, as long as the regulated price falls far below the market clearing price.

Our predictions for the effects of information on supply and enforcement/noncompliance on demand arise from market features that are not unique to kidney transplantation. Shifts in demand with the availability of information can be expected in other industries with significant asymmetric information, such as insurance, pharmaceuticals, and investment banking. Imperfect competition leading to downward sloping demand curves is the rule, not the exception, in essentially all markets. Price controls (or excessive quality) lead to excess demand, the main reason why information alone is not enough. Such controls exist in many markets, especially within healthcare due to the ubiquitous use of Diagnosis-Related Groups in the United States, which pay the same for a given procedure (e.g., kidney transplant, cardiac bypass surgery) regardless of the actual costs incurred. Maybe even more significantly, many countries have national health insurance systems that explicitly control prices and/or facility budgets (e.g., the United Kingdom and Canada), leading to waiting lists for many services beyond transplantation, suggesting that similar dynamics could occur much more broadly. Even in the U.S., Medicaid patients often face similar access issues as in countries with national health insurance systems. In addition, most transplant centers are

located in not-for-profit hospitals, as is common in other countries, and as is common among larger hospitals (greater than one hundred beds) in the U.S.²

In terms of the use and implications of quality measures, transplantation is also not unique. By definition, any payment change from volume of care (traditional fee-for-service payment) to value-based purchasing requires metrics other than units of services delivered. Quality/performance measures are often specifically rewarded, and even when they are not, as in some bundled payment systems, quality needs to be monitored implicitly or explicitly to ensure against unintended consequences. While we study one specific context, the issues we raise (topped out measures, intrinsic competition on quality/performance, gaming of measures) have been shown to arise both theoretically and empirically in healthcare and beyond (Heckman et al., 2011).

Prior U.S.-focused studies in health economics focus on the importance of information availability (through report cards) without enforcement of performance standards. In general, these studies use data on coronary artery bypass graft surgery, a complex medical procedure with comparable short-run mortality rates to those in kidney transplantation, the most common and least complex type of organ transplantation.³ We do not find detectable evidence that report cards affect patient or provider behavior without enforcement, which parallels the results in Dranove et al. (2003), but differs from the results in Cutler et al. (2004) and Kolstad (2013), who find evidence of quality improvements from the

² The 2007 American Hospital Association survey, which includes 231 of the 239 kidney transplant centers operating in 2007 based on the OPTN data, shows that 72 percent of kidney transplant centers are in not-for-profit hospitals, 20 percent in government-owned, and 8 percent in for-profit. This differs from the general sample of hospitals in the 2007 AHA data, but if we restrict the sample of AHA hospitals to just to larger hospitals (over 100 total beds), which includes all but two transplant centers, the non-kidney transplant center ownership types are similar to those in kidney transplantation with 65 percent not-for-profit, 19 percent government-owned, and 16 percent for-profit.

³ Our results are arguably more geographically generalizable than the results in the existing literature. Although heart disease does exceed kidney disease in prevalence, the report card initiatives studied in the existing literature only exist in New York and Pennsylvania (Dranove et al. (2003), Cutler et al. 2004; Wang et al., 2011, and Kolstad, 2013) whereas quality is reported in transplantation for all patients nationwide. This lack of geographic generalizability may be problematic given repeated documentation of variation in access, cost, and efficacy of medical care across the U.S. (e.g., Fisher et al. 2003).

introduction of report cards. Our cream-skimming results for information plus enforcement are similar to results from information alone in Dranove et al. (2003) and our lack of a demand response is similar to results in Kolstad (2013), Schneider and Epstein (1996), Epstein (2006) and Steinbrook (2006). Although enforcement did not exist in these studies, enforcement is increasingly widespread due to the Affordable Care Act, so understanding the costs and benefits arising from enforcement is particularly important, not just from an academic perspective, but from a policy perspective as well.⁴

Papers on healthcare performance measures in other countries have considered these issues in the context of the increasingly common use of quality standards by regulatory entities; how to define quality, gaming behaviors, and risk adjustment in the context of pay-for-performance efforts across a range of medical procedures in the UK, Israel, Australia, Germany, Taiwan, New Zealand, Canada, the Netherlands, and Argentina (Eijkenaar, 2012), in Australia, Canada, United Kingdom, New Zealand, Sweden and the U.S. (Nolte et al., 2011), and in the UK (Bevan and Hood, 2006; Roland and Campbell, 2014). In so far as these papers identify similar concerns and limited evidence of efficacy as shown in the U.S. literature, it seems likely that providers respond similarly to performance standards regardless of level of public provision of healthcare. In some sense, Medicare's ESRD coverage does provide a type of universal access not dissimilar to those found in other countries, although subject to more limitations, such as only three year post-transplant eligibility for coverage, unless covered on the basis of age or disability.

Although we believe our results are generalizable to at least complex medical procedures, if not beyond, kidney transplantation is an important medical procedure in its own right. Kidney disease is the sixth leading cause of death in the U.S.,⁵ some 31 million people in the US have chronic kidney disease, about 600,000 people were living with kidney failure with about 400,000 on dialysis in 2009, rates of diabetes and hypertension (the two leading

⁴ Additional studies of information provision (without enforcement of quality standards) exist in health economics, but are primarily focused on demand side effects (e.g., Werner et al. 2012, Dranove and Sfekas 2008, Bundorf et al. 2009).

⁵ <http://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm>, accessed 05/01/2015.

causes of kidney disease) are increasing, and kidney disease disproportionately affects populations identified with reduced access to healthcare more generally. From a cost perspective, Medicare spent 6.7% of its 2009 expenditures on kidney failure. Because of the substantial maintenance cost reduction and improved quality of life from kidney transplantation vis-à-vis dialysis, maximizing access to kidney transplantation reduces healthcare costs and improves patient welfare.

The paper is organized as follows: Section 2 provides background information on kidney transplantation and the use of survival standards in transplantation; Section 3 discusses the theoretical framework; Section 4, the data and empirical methodology; Section 5, the results; and Section 6 concludes.

2. Background: Kidney Transplantation, Survival Requirements, and Literature

2.1. Overview of Transplantation Process

A patient diagnosed with End Stage Renal Disease (permanent kidney failure) has two options, dialysis or a kidney transplant. Costs are lower and the quality of life is higher in the long run for patients who receive transplants making this the preferred option (Matas and Schnitzler, 2003). Most patients undergo dialysis prior to receiving a transplant due to the extensive waitlist for deceased donor kidneys and the limited number of living donors that are a close biological match, willing to donate, and healthy enough to donate. The patient's doctor or the dialysis center generally refers the patient for transplantation. Usually, patients must choose a transplant center within their insurer's preferred provider network and these options may be quite limited, depending on the insurer. For example, in the last six-month period in our sample, from 01/01/2011 and 06/30/2011, 89 of the 219 transplant centers operating in that period performed fewer than twenty transplants, falling short of Aetna's annual volume requirement of forty transplants per year.⁶

⁶ <http://www.aetna.com/healthcare-professionals/documents-forms/Aetna-Institutes-of-Excellence.pdf> Accessed 06/18/14.

The decision to waitlist a patient or remove a patient from a waitlist lies with the transplant center, creating the possibility of strategic responses to regulatory incentives. Waitlisting criteria include the ability to pay for the transplant and long-term immunosuppression medications after the transplant, proof that the patient can arrive at the hospital within the limited number of hours in which a kidney can be used for transplant, and no physical or psychological contraindications. Once on the waitlist, the patient must undergo frequent testing to maintain his or her priority status. The patient then waits to be matched with a donated organ.⁷

Deceased donor organ allocation occurs exclusively through the OPTN, created by Congress with the National Organ Transplant Act in 1984 and operated ever since by the United Network for Organ Sharing, a nonprofit contractor. Since 2008, the OPTN has had authority over live organ donation as well. All live donor organ procurement must now occur in one of the OPTN's approved hospitals, all live donors must undergo an extensive screening process prior to donation, and live donor follow-up data must be collected for at least two years post-donation. The OPTN oversees of fifty-eight local Organ Procurement Organizations, which each, in turn, are responsible for a single exclusive Donation Service Area, containing between one and thirteen kidney transplant centers.

When a hospitalized patient dies or death is imminent, the hospital must notify the local Organ Procurement Organization, as required by law since 1998. Personnel from the Organ Procurement Organization contact the family of the patient to obtain consent, unless donation already was authorized. If consent is given, then the United Network for Organ Sharing follows the kidney allocation algorithm to determine the best matches for the organ, based on medical criteria such as blood type, organ size, and disease severity. Time

⁷ Although most patients are waitlisted at some point during the transplant process, living donors can direct their donation to an individual patient, allowing that patient to opt out of the deceased donor waitlist. Directed deceased donation is possible, but it is not tracked in the data and rarely occurs according to conversations with staff at a major U.S. transplant center. Undirected live donor organs are also extremely rare and such organs are allocated by the OPTN in the same manner as undirected deceased donor organs.

on the waitlist serves as a tie-breaker. In general, kidneys are first allocated within a Donation Service Area, before being offered regionally, and then nationally.

The allocation process is crucial to our analysis because it limits a transplant center's opportunities for control over transplant recipient selection. Although a center can choose whom it wants to waitlist and whom it wants to remove from their waitlist, it does not control the prioritization of patients on its waitlist in the deceased donor allocation process. In other words, it cannot "cherry pick" from its waitlist according to which organs become available or in order to delay transplanting a high risk patient. The only solution for a center to avoid having high risk patients negatively impact its survival rates is to not list such patients or to remove such patients from their waitlist before they are matched with an organ. All such removals must be recorded in the data submitted to the OPTN.

A patient with a live donor does not have to participate in the waitlist process, but usually does so in case the live donor decides not to donate for medical or personal reasons. Most live donors are family or friends and live donor kidney transplants do have better outcomes, all else equal, because the timing of the transplant can be optimized based on the health of the patient and the donor. (However, some transplant centers will provide a transplant to patients deemed too sick for a deceased donor transplant if they have a live donor available.)

2.2. Quality Standards and Reporting in Transplantation

Since 1991, the OPTN has used quality standards to regulate which centers are permitted to access the national organ supply and perform organ transplants. In 1991, the 5% of centers with the lowest organ graft and patient survival were flagged for review by OPTN. Today survival standards are based on deviations from risk-adjusted expected survival rates based on national averages. Center-specific performance measures and compliance status have been publicly available since 1991, but their accessibility increased substantially in 2001 when online reporting replaced paper-only reporting. Enforcement of these policies also has changed over time. Prior to July 2007, the OPTN was the primary oversight organization but only had twice recommended to the Department of Health and Human Services that a

transplant center should no longer be permitted to perform transplants. CMS became concerned that, although centers were required to meet certain standards to begin performing transplants, oversight did not really exist thereafter, likely in part due to a lack of any serious penalties from not meeting the OPTN's minimum volume and survival standards. As stated in the Final Rule establishing the increase in CMS oversight:

“The OPTN generally takes a collegial approach and assists centers in improving their performance, while we generally take a regulatory approach which sometimes may lead to termination...” (CMS, 2007)

A proposal for CMS's new enforcement role was provided for public comment in 02/04/2005, was approved 03/30/2007, and became effective 06/28/2007. The enforcement process proceeds as follows. A center is flagged for noncompliance based on its failure to meet expected one year post-transplant survival rates for patient and/or graft survival. (All patient deaths are included as graft failures even if the death is not caused by failure of the organ transplant and it is possible for a center to be compliant with one measure and not the other.) The specific threshold for noncompliance is based on the difference between observed patient deaths (or graft failures) within one year of transplant and expected patient deaths (or graft failures) within one year of transplant based on national averages as calculated by SRTR using Cox proportional hazard models. These are calculated, separately for adult and pediatric transplant programs, for all types of organ transplants except for heart/lung transplants and pancreas transplants. In other words, a hospital might have an adult kidney program out of compliance and a pediatric heart transplant program in compliance during the same period. For simplicity and clarity, we focus only on compliance with regard to adult patient deaths after kidney transplantation. (Pediatric patients tend to suffer from different diagnoses and constitute a much smaller sample than adult patients.) Noncompliance with either patient deaths or graft failures can lead to audit and possible closure. In order to be compliant, a center must meet at least one of the following survival standards for patient deaths (or graft failures):

1. Observed – Expected ≤ 3
2. Observed/Expected ≤ 1.5
3. The one-sided p -value for the difference between Observed and Expected ≥ 0.05

If flagged for noncompliance, a center may file a “request for consideration of mitigating factors”⁸ in order to attempt to avoid the next step, an extensive (and expensive) audit conducted in conjunction with CMS to determine why the center’s survival outcomes did not meet expectations and how outcomes will be improved going forward. If the center fails the audit, then CMS has the authority to close the center and prevent it from performing further transplants. A request for consideration of mitigating factors does not guarantee that a center will avoid penalties, but could delay the center’s response to being in noncompliance if the center waits for the outcome of its request before changing its behavior. In general, a center has 210 days to improve quality or have a request for consideration of mitigating factors reviewed by CMS before termination. Between 07/01/2001 and 02/10/2009, CMS reviewed twenty-two requests for consideration of mitigating factors. Most centers that had failed to comply with the CMS survival standards at that time filed requests for consideration of mitigating factors, but only nine were approved, primarily based on evidence of improvement in shorter term survival outcomes not yet evident in the data used for establishing compliance by CMS (Hamilton, 2009). Although survival requirements exist for all types of organ transplants, we focus on kidney transplants for a variety of reasons. They are the most common type of transplant with 467,620 transplant candidates waitlisted and 375,406 transplanted between 10/01/1987 and 03/14/2014. On 03/14/2014, 102,869 individuals were waiting for a transplant and in 2013, 17,657 patients received a transplant.⁹ Figure 1 shows the number of waitlist additions, removals and transplants by period.

⁸ <http://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/CertificationandCompliance/Transplant.html>

⁹ <http://optn.transplant.hrsa.gov/latestData/advancedData.asp>. Accessed 06/18/14.

The primary enforcer of survival standards, CMS, is significantly more involved in kidney versus other types of organ transplants due to the Medicare entitlement for individuals with End Stage Renal Disease. All individuals qualify regardless of age if they paid Medicare taxes for at least ten years or had a spouse who did so (or a parent in the case of a minor). Medicare coverage becomes effective once a person is waitlisted, but is secondary to any employer insurance for the first thirty months after diagnosis. Medicare covers both live and deceased donor transplants with a coinsurance rate of 20% (Medicare Coverage of Dialysis and Kidney Transplant Benefits, 2014). In our data, 44% report Medicare as their primary insurance provider and 15% report Medicare as their secondary insurance provider.

Due to the existence of dialysis, kidney transplant centers arguably have more flexibility than other types of organ transplant centers in removing patients from the waitlist; doing so will not necessarily result in death as it would for other types of transplants. It also is more obvious whom to remove from the waitlist because patients who are sicker prior to transplant tend to have worse outcomes post-transplant. (This is not the case for all types of transplants. For example, “sicker in” does not necessarily mean “sicker out” for liver transplants.) Kidney transplant candidates (individuals waitlisted for a kidney transplant) also have somewhat more control over where they waitlist, relative to other organ transplant candidates because more kidney transplant centers exist for any other type of organ and Medicare’s minimum annual volume requirement of three kidney transplants per year is almost universally satisfied. This also implies that in general, centers do not have to maintain kidney transplant volumes while simultaneously meeting or improving survival standards. Other insurers, however, have different minimum volume standards. (As mentioned earlier, Aetna requires forty adult kidney transplants per year in order for a center to remain part of its network.)

The literature on the effect of survival requirements in transplantation prior to the increase in CMS oversight is limited to Howard and Kaplan (2006), who study patients waitlisted for kidney transplants between September 1999 and October 2002 and find no effect on

demand from the posting of survival outcomes online through the Scientific Registry of Transplant Recipients, which began in July 2001.

The increase in CMS oversight in July 2007 generated interest in transplantation and health economics. Editorials appeared in major transplantation journals discussing potential negative effects of these standards on patient selection and research activity (Abecassis et al., 2009; Roberts, 2012). A survey of sixty-three participants at the Transplant Management Forum in 2009 gave suggestive evidence that transplant centers increase patient and donor selectivity and change clinical protocols in response to the increase in oversight in July 2007 (Schold et al., 2010). Both the Schold et al. (2010) survey and a paper by Hamilton (2009) provide evidence that CMS requirements are enforced. The survey responses indicated that thirty-one percent lost insurance contracts as a result of not meeting quality standards. Two recent working papers find an increase in donor selectivity (Cox et al., 2012) and an increase in wait time and the average health of patients among those receiving transplants between June 2007 and 2010 (Schnier et al., 2014) at centers that are not in compliance with the CMS quality standards for kidney transplants.

These articles on survival standards in transplantation suggest that the most recent regulatory changes in 2007 have led to some changes in the behavior of transplant centers and/or their patients. Discussions with transplant personnel at a major center yielded similar assessments. Publicly available information on the full extent of the effect on patient outcomes is not available beyond comparisons of means in a report issued by the University of Michigan Kidney Epidemiology and Cost Center (2011) and the effects studied in the working papers by Cox et al. (2012) on donor selectivity and by Schnier et al. (2014) on wait time and the average health of patients receiving transplants. No published sources empirically assess other patient outcomes or how and when centers respond to approaching or breaching the survival standards.

3. Theoretical Framework: The Market for Kidney Transplants

Factors affecting the marginal cost of transplantation, such as the enforcement provisions instituted by CMS, can affect the supply of transplants, primarily due to the lack of a compensating increase in price. Figure 2 depicts a simple demand-and-supply framework for a typical transplant center. O_S represents the total supply of organs. The marginal cost is upward sloping due to variation in the quality of the underlying organs and patients. Lower quality organs and sicker patients are associated with a higher rate of complications, shorter post-transplant survival periods, and longer surgeries. We know the regulated price must be sufficiently low so that not all organs are used (Tanriover et al. 2014), leading us to a quantity supplied of Q_S .

Transplants, as with other many medical procedures, are subject to Medicare's Diagnosis-Related Group (DRG) pricing system. A hospital receives a fixed payment for a kidney transplant, regardless of the actual costs associated with performing the procedure, thus providing a disincentive to treat higher cost patients.^{10,11} At the regulated DRG price and with the current benefit/cost structure faced by patients deciding to pursue a transplant rather than kidney dialysis and the existing donor availability, which is limited by the prohibition of compensation for living or deceased donation, demand far exceeds supply with a kidney transplant waitlist of over 100,000 candidates. In 2014, just over 17,000 kidney transplants were performed.¹² (Figure 1 clearly depicts the disparity between supply and demand.) Demand is downward sloping and somewhat price elastic due to insurer involvement, the high price of transplants, and patient deductibles and coinsurance rates (20% for physician services and immunosuppression medications with Medicare). With quality reporting, demand could shift up to D_G or down to D_B depending on how the quality information affects the perceived marginal benefit of a transplant at the center. Given the

¹⁰ The DRG system does allow for outlier payments, but only if the costs are so high that they pass a certain threshold well above the underlying DRG. Transplant centers are not reimbursed for the costs between the DRG and the "outlier" threshold, only receiving compensation for costs above the outlier threshold.

¹¹ Although other insurers cover kidney transplants, Medicare's significant involvement drives the price, so we use Medicare's pricing system to develop our model. Our model's predictions should hold even with the inclusion of other payers. Due to a scarcity of information on other insurers' transplant reimbursement practices and possible price-taking behavior on the part of transplant centers, we abstract away from these issues in our simple model.

¹² <http://optn.transplant.hrsa.gov>, Data Reports, accessed 4/10/2015.

large and persistent shortage in the current marketplace, it seems reasonable to assume that information would never shift demand in far enough for it to affect the quantity supplied.

In such a context, we are able to abstract away from any demand influences. Although the provision of information may shift firm-level demand curves by increasing demand at high quality centers and decreasing demand at low quality centers, the extreme excess demand in the market means that the center's total demand will still continue to exceed the donor supply, as long as the price is fixed far below equilibrium. In other words, the provision of information will not affect the supply of transplants by any type of center because even for low quality centers, demand significantly exceeds supply due to DRG-based payments and the limited supply of donor organs. (Even without massive excess demand, the lack of a demand effect may not be unique to kidney transplants. Evidence exists that quality report cards do not affect cardiologist and managed care insurer referrals (Schneider and Epstein, 1996; Mukamel et al., 2000; Epstein 2006, Steinbrook 2006).)

This leads us to our first prediction – transplant centers will not respond to the provision of information alone. Why would they given that all of them have excess demand regardless of quality? In other words, information will not lead to improvement in quality or trigger gaming behavior.

The remaining questions we seek to answer arise from the predicted effects of enforcement and noncompliance on supply. Enforcement alone, even for compliant centers, shifts up the marginal cost of performing transplants by amplifying the potential risk of performing transplants and increasing costs due to additional quality controls implemented to avoid noncompliance. As seen in the graph, the difference in the marginal cost of treating healthy patients using good organs is similar across regulatory regimes because these patients pose a very low risk to transplant centers. As centers increase the number of transplants performed, the marginal cost increases more rapidly in the period with enforcement/noncompliance because the regulatory risk increases the marginal cost of

treating sicker patients using worse organs, i.e., the supply curves become more inelastic. This shifts the quantity supplied from Q_S to Q_S' . If a center is noncompliant, the marginal cost of performing an additional transplant increases still further, potentially reducing the quantity supplied even more. However, this result assumes that centers cannot reduce the expected marginal cost of performing an additional transplant by decreasing the riskiness of the patients receiving transplants, thus shifting the new supply curve down towards the original supply curve. The effects on supply lead us to our hypotheses regarding center behavior in the presence of the new CMS rules. We expect all centers to compensate for the increased marginal cost created by the risk of audit and closure by reducing the average risk of potential transplant recipients (i.e., those individuals on the waitlist). Those centers flagged for noncompliance will do so to an even greater extent.

4. Data

Center-identified data on waitlisted and transplanted patients were obtained under a confidentiality agreement from the United Network for Organ Sharing. These data cover the period from 10/01/1987 through 09/30/2013.

Center-level data on expected survival rates and compliance with survival standards were obtained from the Scientific Registry for Transplant Recipients. These data cover patients transplanted between 07/01/1997 and 06/30/2011. Reports are issued on a biannual basis every January and July and are based on a rolling 2.5 year cohort of patients, with the first cohort with on-line reporting spanning from 07/01/1997 to 12/31/2000.

We organize the data into six month periods at the center level. The data cover 23 cohorts for which SRTR issued reports (starting in 2001), 272 transplant centers, and 271,555 waitlist removals including 160,614 transplants. This yields 3,875 center-period observations for transplants (with a one year follow-up period) and 4,482 center-period observations for waitlist removals. Figure 3 illustrates how the data are organized. The column headings 1 to 15 are sample six month periods. Each cohort spans five six month periods as shown by the lightest shade of gray. A one year lag exists between the end of a cohort and the issuing of the SRTR report because the report is based on one year post-transplant survival rates. The

“R” designates the period at the beginning of which the report is issued. We anticipate most effects to occur in patient populations transplanted or removed from waitlists in the six month period at the beginning of which the SRTR report is issued. The darkest shade of gray indicates where we expect to see the largest effect on patient outcomes and the medium shade of gray indicates a sustained effect extending beyond the periods directly after a report becomes publicly available. Quality (survival) improvements may take longer to become evident, so we extend our analysis into the periods a year after the report was issued as depicted by the second set of light gray boxes. Figure 4 shows the percent of centers, which are flagged for noncompliance in each period. On average 10% of centers are flagged for noncompliance with a maximum of 12.5% in January 2007 and a minimum of 7% in July 2011. One hundred and two centers are flagged at some point during the study period. Sixty-four percent of flags for noncompliance are followed by a flag in the subsequent period.

In the Table 1, we detail entry (no evidence of transplants performed in prior years), shutdown (at least one period after entry with no transplants performed), re-entry (re-opening after a shutdown period) and exit (no evidence in our sample of re-entry). Re-entry can only occur after shutdown. Exit is permanent for the period of our sample, but may underestimate the long-run probability of re-entry in that we do not follow the centers beyond the end of our sample period. Focusing on the period in which public reporting was available (beginning July 2001), 15 centers enter, 34 shut down (i.e., have performed transplants previously, but do not perform transplants for at least one period), 23 of those that shut down exit and 11 re-enter later in the sample period. Of those that exit, 5 were deemed noncompliant that period and 13 had been flagged within the prior three periods. For shutdown, the numbers are 7 and 15. On average, centers that shut down have received 1.4 prior flags (which may or may not be consecutive) and centers that exit have received 2 flags on average. Centers that re-enter only have 0.2 flags on average. Although attrition bias could be a concern, given these numbers, we note that the majority of centers that shutdown and just under a majority of centers that exit were not flagged for noncompliance within the prior three periods. In terms of prevalence of shutdown and exit among

noncompliant centers, of the 101 centers that are flagged in the sample, 85 do not shut down and 89 do not exit. Therefore, noncompliance does not lead to exit in general.

Additional information on transitions is provided in Tables A1 to A4. Table A1 shows entry, shutdown, exit and re-entry by period. Table A2 presents the period and number of flags for each of the twenty-three centers that exit the data. Table A3 shows that the total number of flags received by a center varies from zero to fourteen, with the majority of centers receiving no flags and the majority of flagged centers receiving three flags or less during our sample period. Table A4 provides information on consecutive flags. Although most flags are not consecutive, not an insignificant number are, so we address this issue in our robustness checks.

4. Empirical Methodology

We use the following regression model to test the predictions of our model

$$Outcome_{ht} = \alpha_0 + \alpha_1 Flag_{ht} + \alpha_2 PostCMS_t + \alpha_3 Flag_{ht} * PostCMS_t + \alpha_4 Trend + \alpha_h + \varepsilon_{ht}$$

The subscript t specifies the period in which an SRTR report is issued, with the first report ($t=0$) issued on 07/01/2001 and the last period ($t=23$) corresponding to the report issued on 06/30/2012. An individual transplant center is referenced by the subscript h .

$Outcome_{ht}$ includes the following center six-month period level outcomes: the one year post-transplant survival rate, the waitlist transplant rate, the waitlist deceased donor transplant rate, the waitlist live donor transplant rate, the waitlist removal rate of patients deemed too sick for transplant, and the rate at which waitlisted patients are transferred to

other centers. We use center-period level rate outcomes rather than patient-level outcomes to diminish the influence of outliers, to avoid larger centers driving the results, and to evaluate the data at the level at which the regulation is set, i.e., at the center-period level. (We run negative binomial regressions using the count of the outcomes as a robustness check.¹³)

Table 2 shows descriptive statistics for the outcome variables, both counts and rates. Transplantation is not a high frequency procedure; the average number of transplants performed at a center in a six-month period is thirty-four with a standard deviation of twenty-nine. The three post-transplant survival rates tracked in the online SRTR reports are one month post-transplant survival, one year post-transplant survival, and three year post-transplant survival. Only one year post-transplant survival is used by CMS in its survival requirements. Average survival rates are ninety-eight percent, ninety-one percent, and sixty-eight percent respectively.

The next set of variables in Table 2 includes total waitlist removals and the documented reasons for removal. On average 57% of patients removed from the waitlist are removed in order to receive a transplant with 41% receiving deceased donor transplants and 15% receiving live donor transplants. The other removal reasons studied in this paper include “too sick for transplant” (5%) and transferred to another transplant center (13%). The only common removal reason we do not study is the waitlist death rate at 15%.¹⁴ Uncommon removal reasons include, for example, patients who refused transplants, whose condition improved so that a transplant was no longer needed, candidates removed by mistake, patients transplanted in other countries, and candidates that could not be contacted.

¹³ Patient-level regression analysis using a linear probability or logit model yields the same pattern of results but clearly weights larger centers more heavily.

¹⁴ Deaths on the waitlist neither benefit nor hurt transplant centers in terms of compliance and all coefficients except the main effect on the *PostCMS* variable are statistically insignificant in all specifications.

$Flag_{ht}$ equals one if a transplant center was noncompliant, i.e., did not meet all three survival requirements in the report issued at time t , with the effect tested on patients removed from the waitlist or transplanted at time t , $t+1$, $t+2$, and $t+3$.

$PostCMS_t$ equals one if the outcome is measured in a period after CMS increased its oversight of transplant centers. Of the twenty-three six month periods in our sample, eight periods occur after CMS stepped up oversight.

The other three terms include a linear time trend measured at the six month period level (*Trend*), center fixed effects (α_h) and an error term (ε_{ht}) with clustering at the center level to adjust for heteroskedasticity and serial correlation. We use a trend rather than period-level dummies in order to better identify the effect of the *PostCMS* variable. The results, however, are similar with either time control specification for the outcomes studied. In general, as shown in Figure 5, survival rates have been fairly constant with some increases in survival shortly after CMS began enforcement. There were no other abrupt changes in the regulation of kidney transplants or changes in the kidney transplantation procedure itself. (Some gradual shifts in the types of immunosuppression used during this period did occur as did increasing usage of laparoscopic surgical techniques for live donors.)

5. Results

Post-Transplant Survival

We begin by studying whether or not being flagged as noncompliant with survival standards leads to improvements in one year post-transplant survival outcomes, the goal of the survival requirements. Survival is measured on the basis of when a transplant was performed. In other words, does the one year post-transplant survival rate for patients transplanted in time t increase or decrease? Table 3 shows the results for one year post-transplant survival, the measure used by the OPTN and CMS to determine whether or not a

center is noncompliant. As explained earlier, centers are flagged at the beginning of period t for performance during the 2.5 year cohort ending at the end of period $t-3$. The survival rate is evaluated for each six month period after the SRTR report is issued as denoted by the column headings in Table 3. In order to evaluate whether some time is required for transplant centers to change their behavior and to determine if the effect of being flagged for noncompliance persists for the longer term, we also evaluate outcomes at time $t+1$, $t+2$, and $t+3$. Table 3 shows no statistically significant evidence of improvements in one year post-transplant survival outcomes from being flagged for noncompliance, either before or after CMS stepped up oversight, and shows no effect from enforcement alone.¹⁵ The lack of an effect also is not due to a change in the patient population, as evidenced in patient- and center-level regressions controlling for a wide range of risk adjustment factors used by the SRTR to calculate expected survival rates.¹⁶ (See Table A5 for center-level regressions including the SRTR risk adjustment variables.)

Gaming Behavior

Although we find no evidence of an improvement in survival outcomes from these regulations, it may be that they still affect transplant center behavior in other ways. More specifically, we evaluate effects on waitlist management with the hypothesis, supported by the survey evidence in Schold et al. (2010), that deceased and live donor transplant rates will fall and the rate of patients deemed too sick for transplant or transferred to other centers will increase. We use the same approach as with survival except that the denominator is the total number of waitlist removals rather than the total number of transplants. Tables 4 through 8 show the effects of noncompliance on these variables for the four periods after noncompliance with patient survival requirements has been documented by the SRTR reporting system. All statistically significant effects exist *only* after

¹⁵ Regressions evaluating survival outcomes in periods subsequent to $t+3$ also show no evidence of improvements in survival.

¹⁶ Results available upon request. Risk adjustment variables include gender, race, blood type, wait time, diagnosis, panel reactive antibodies, previous transplant recipient, age, obesity, type of insurance, cold ischemia time, deceased non-heart-beating donor, donor age, donor with a cerebrovascular cause of death, diabetic donor, hypertensive donor, donor organ pumped, donor race, donor creatinine, donor-recipient weight ratio, extended criteria donor, number of HLA mismatches, donor gender, and organ share type (local, regional or national).

CMS increased enforcement in 2007. The effects are most statistically significant for the transplantation outcomes and document a 7.4 percentage point (13% relative to the mean rate) decrease in the total transplant rate at time t , with a 4.9 percentage point (12% relative to the mean rate) reduction for deceased donor transplants and a 2.5 percentage point (17% relative to the mean rate) reduction for live donor transplants. For the overall transplant rate and the deceased donor transplant rate, these effects persist through $t+3$.

If a smaller percentage of patients removed from the waitlist are removed in order to receive a transplant, then transplant centers must be removing these patients for other reasons. We test whether the decrease in the transplant rate is mirrored by an increase in other outcomes, specifically the rate at which patients are deemed too sick for transplantation or are transferred to another center. While the transplant rate decreased by about seven percentage points, the results in Tables 7 and 8 show that the number of patients deemed too sick for transplant increased by about almost two percentage points and the number of transfers increased by about three percentage points, i.e., these two outcomes account for the majority of the decrease in the transplant rate. Put another way, our results suggest that a large percentage of patients who would otherwise have received a transplant are now being deemed too sick for transplant or are being transferred (or transferring) to other centers.

From a patient perspective, a decreased rate of transplantation and an increase in patients removed as too sick for transplant or transferred to other centers cannot be strictly welfare-improving for that patient. (Given the lack of an improvement in survival overall, welfare benefits due to an increase in total generated life years at the societal level also do not appear to exist.) Not receiving a transplant is associated with decreased quality-adjusted life years while on dialysis and a decreased overall life expectancy. The only possible welfare-improvement would arise if patients have a higher survival probability at the center to which they transfer net of the costs and risks of the actual transfer process. In addition, the likelihood that transfers are driven by patients choosing to change centers is extremely small because all effects from noncompliance appear tied to CMS enforcement rather than

to a correction of information asymmetries between patients and centers through the public reporting of survival outcomes that existed throughout the study period. Put another way, public reporting alone does not affect waitlist outcomes, so it must be enforcement driving the changes and enforcement effects are more likely to arise from changes in transplant center behavior rather than shifts in patient behavior given the lack of any additional information and the fact that patients may be even more complacent in choosing transplant centers if they feel they can rely on CMS to ensure a minimum quality level.

The results for patients delisted as too sick for transplant are quite large in magnitude. With a mean of 5% removed as too sick for transplant in a given six month period, the coefficient of 0.018 is associated with a statistically significant increase of 36% in the rate at which patients are deemed too sick for transplant. Although only marginally significant, the pattern of results for transfers indicates approximately a 24% increase in transfers from the mean of 13%. These results persist using a negative binomial specification controlling for the number of waitlist removals or if we control for patient case mix.

Based on the coefficient on the *PostCMS* variable, the existence of enforcement alone is insufficient to induce a strategic response from transplant centers beyond some evidence of a reduction in the number of transplants, especially from deceased donors.

In regressions not reported in this paper (available upon request), we do find a reduction in the number of patients waitlisted at centers flagged for non-compliance, but this is likely at best a secondary effect given the lack of theoretical predictions or empirical evidence of a demand effect. Due to the importance of wait time during this period and with many patients waiting years for a transplant, decreasing the number of sick patients on the waitlist is unlikely to yield risk reduction benefits except in the long-run. The average patient waited two years on the waitlist before removal in 2007 and that number had increased to 2.4 years by 2012. Shifts in the waitlist demographics may allow for long-run cream-skimming, but are unlikely to have an immediate impact on compliance. Schneir et al. (2013)

document an increase in waiting time post-noncompliance, but the mechanism through which this occurs is not clear.

A more immediate waitlisting outcome we do identify is a reduction in the number of waitlisted patients willing to accept an Extended Criteria Donor (older with higher rates of hypertension and elevated creatinine levels), as determined by the waitlist modification data. We do not find a statistically significant effect on waitlist temporary inactivations or shifts in the initial status (active, inactive, urgent, or critical) at noncompliant centers or from enforcement absent noncompliance. We do find a slight reduction in the rate of patients waitlisted with diabetes, a factor associated with worse post-transplant outcomes relatively to other diagnoses.

Preemptive Responses

Every SRTR report lists the number of observed and expected deaths, allowing transplant centers (and patients) to review how well a center is performing relative to each of the sub-measures used to determine noncompliance. Therefore, the transplant center behavioral response we see from actual noncompliance may underestimate the effect of the survival requirements if centers respond to approaching noncompliance in a similar manner to how they respond to actual noncompliance. We focus on the “Observed – Expected” outcome in order to evaluate whether centers respond before actual noncompliance.

Because centers are unlikely to respond linearly to approaching the threshold for noncompliance and because some insurers require that the observed be no greater than the expected (although CMS allows a difference of three), we create categorical variables for whether a center has observed deaths equal to or less than expected deaths, whether the observed deaths are greater than the expected deaths but still below the CMS threshold, and whether the number of observed minus expected deaths exceeds the CMS threshold of three. This last group includes centers flagged for noncompliance due to not meeting all three survival thresholds and centers that are still compliant although they failed to meet the “Observed – Expected ≤ 3 ” threshold. As shown in Table 9, about forty-percent of

centers are non-compliant with this threshold at some point during the sample period. In terms of center-period level observations, 57% of observations have an observed number of deaths no greater than the expected number of deaths, 31% have observed deaths greater than expected deaths but still within the acceptable range for compliance with CMS' standards, and 12% of center-period observations have observed deaths that exceed expected deaths by more than three. Two percent of center-period observations fail to meet this specific survival threshold, but are still overall compliant with the CMS standards. (Omitting this two percent of centers does not materially affect the results in Table 10.)

Table 10 provides results for the full sample and shows that centers do engage in preemptive behavior in terms of modifying their waitlist management when they approach noncompliance but before they are actually noncompliant and that simply comparing noncompliant periods with compliant periods would underestimate the magnitude of the effect of the CMS requirements. All outcomes are measured for patients receiving transplants or removed from the waitlist during the period at the beginning of which the SRTR report was issued (time t). As with full noncompliance with all three survival requirements, we see no effect on one year post-transplant survival.

For the delisting outcomes, we see similar patterns to what we saw with the overall results from noncompliance; a decrease in deceased and live donor transplant rates and increases in the rate at which patients are deemed too sick for transplant or transferred to other centers. The results in Table 10 also document a preemptive effort on the part of centers to avoid noncompliance. For overall transplant rates and deceased donor transplant rates, the effect of approaching the threshold leads to a statistically significant reduction in the transplant rate while the effect of exceeding the threshold is about twice as large in magnitude. For live donor transplants, the difference is not quite as stark, but centers respond to almost breaching the threshold and even more strongly to actually breaching the threshold. For too sick for transplant, centers appear to wait until actual noncompliance with the measure rather than preemptively kicking patients off the waitlist. Transfer rates are statistically significantly different for centers for which observed patient deaths exceed

expected patient deaths but the effects of being near noncompliance and noncompliance are similar in magnitude and not statistically significantly different from each other.¹⁷

The statistically significant reduction in the transplant rates and the increase in the transfer rate are likely associated with greater caution on the part of transplant centers post-CMS enforcement, in an effort to decrease the marginal cost by decreasing the number of risky transplants performed. As shown in the Figure 1, transplant rates appear to plateau around the time of CMS enforcement likely due in part to the fixed constraint in terms of the number and quality of donor organs available. We believe that the lack of a statistically significant preemptive effect for removals of patients deemed too sick for transplant is due to the greater emotional impact and possible reputation damage from kicking sick patients off the waitlist.

5.1. Robustness Checks

We run a variety of robustness checks including using counts rather than rates as outcome variables, one month and three year post-transplant survival outcomes rather than one year post-transplant survival outcomes, and noncompliance with graft failures rather patient deaths. We also evaluate subsamples of noncompliant centers: only those flagged for the first-time, omitting those flagged for the first time, and omitting centers flagged more than three times during our sample period. In addition, we test whether the most sophisticated and/or wealthy patients (those that simultaneously waitlist at multiple transplant centers) exhibit a demand response even if we find no overall evidence of a demand effect from information. Due to the broad literature on the volume-outcome relationship in healthcare, we also control for volume in a variety of ways. In all cases, these checks support our main results.

Because the outcomes we are evaluating are rates, it could be that improvements in survival do exist, but that centers also have increased the number of transplants and extended them to patients with a lower probability of survival. Although unlikely, we consider this

¹⁷ Regressions excluding centers that are overall compliant but fail to meet the “Observed – Expected ≤ 3 ” threshold or categorizing those centers with centers in the “ $0 \leq \text{Observed} - \text{Expected} \leq 3$ ” in an “almost noncompliant” group yields similarly interpretable results.

possibility in Appendix Table A6. After CMS stepped up oversight, we see a large and statistically significant *reduction* in the number of transplants at centers flagged for noncompliance. In other words, despite a possible improvement in the pool of patients receiving transplants through more cautious waitlisting practices, the number of survivors per one hundred transplants remains statistically unchanged. (Due to the count outcome with evidence of statistically significant overdispersion, Table A6 uses a negative binomial specification.) In addition, we verified the lack of an effect using a negative binomial specification with the number of survivors as the outcome variable and the number of transplants as an independent variable as reported in Table A7. (Similar count regressions confirm our waitlist removal regression results.) We confirm in Table A8 and A9 that, in general, no improvement exists for one month and three year survival periods, the two additional measures reported by the SRTR but not used by CMS for measuring noncompliance. We do see an increase in three year post-transplant survival among noncompliant centers from information alone. Given that this is the only time any evidence of survival improvement exists in the overall sample, we believe that this result may be spurious. In addition, we only observe three year survival periods for patients transplanted prior to July 1, 2009, leading to a limited number of observations, particularly in the CMS enforcement period. On the other hand, this result may provide some hope that alternative survival measures might function better than one year post-transplant survival.

Although we anticipate the greatest impact on center and patient behavior from noncompliance with patient death requirements, we test whether a similar effect exists for graft failures and whether flagging for both types of noncompliance has a differential effect as well. A failure to comply with graft survival requirements leads to the same response by CMS in terms of auditing and potential closure, but carries less of a burden from a public relations perspective because such noncompliance is not as easily accessible on the SRTR website, graft failures are not necessarily fatal, and centers may be better able to more precisely change practices leading to graft failures, requiring a less dramatic response than patient survival noncompliance from a waitlist management perspective. As shown in Appendix Table A10, about 10% of centers are flagged for noncompliance with patient

survival requirements and about 11% are flagged for noncompliance with graft survival requirements. Approximately 5% are flagged for both. Seventy-three centers are ever flagged for both and the correlation between the two types of flags is 0.42.

As shown in the first two columns of Appendix Table A11, the effects from flagging for graft failures are quite similar to those from noncompliance with patient survival requirements. Centers that are flagged for both graft failures and patient deaths exhibit the greatest change for the overall transplant rate, the deceased donor transplant rate, and the transfer rate. None of the three types of flags including in Table A11 indicate any effect from regulatory noncompliance on one year post-transplant survival.

We now explore subsamples based on flagging incidence. It is not clear whether first time flags should generate a larger or smaller effect on center behavior relative to consecutive flags. On the one hand, centers flagged for the first time may react more strongly in order to maintain their reputation. On the other hand, centers flagged repeatedly are less likely to have simply had a random event that resulted in noncompliance. In the end, we are limited by our data. (The flagging incidence is shown in Figure 4 and Tables 1, A3, and A4 provide additional statistics on flagging, including frequency of flagging and total rates of flagging.)

Sample size restrictions make it difficult to evaluate the effect of repeated flagging in consecutive periods, and even just considering first time flags leads to a significantly reduced sample. In other words, even the regression results considering first time flagging effects should be interpreted with caution. The results in Appendix Table A12 compare observations before a center is ever flagged with observations in the period in which a center is first flagged. This leads to a loss of 932 observations for center-periods associated with repeated flags for noncompliance. The same pattern exists in these results as in the main results although the magnitude and statistical significance is somewhat diminished as one would expect given that consecutive flags are likely to be met with greater regulatory and demand-side responses than first-time flags and hence prompt a larger response on the part of the transplant centers as well. In addition, we are attempting to identify off a much

smaller number of observations. Only 2.5% of centers are flagged for the first time in the average period in which an SRTR report is issued and this drops to less than 1.5% in the period after CMS increased oversight because many centers already had been flagged for the first time. Overall, 10% of centers are flagged in a given period.

We ran additional regressions omitting centers that were flagged only once (Table A13) or flagged more than three times (Table A14) in the sample and find that the results generally hold, although centers that are flagged multiple times appear to have a larger response. This may be driven by the fact that response is only evident post-CMS enforcement, which is later in the period, and therefore, centers are more likely to have been flagged multiple times.

In the Table 1, we detailed entry, shutdown, exit, and re-entry. Although most noncompliant centers do not shutdown or exit within our sample period, we reran our main regressions including only those centers that are noncompliant but do not shutdown or exit during the sample period (Table A15). Compared with the main results, centers that do not shut down are more likely to decrease their transplant rate by decreasing live donor transplants rather than cadaveric donor transplants. We find no evidence of effects on patients removed as too sick for transplant or transferred away from noncompliant centers. However, it appears that enforcement without noncompliance may be inducing a preemptive effect given the significance of the *PostCMS* variable for cadaveric donor transplants and for removals as too sick for transplant.

The significance of the *PostCMS* variable led us to rerun the regressions seeking to identify a preemptive response and we find that, although noncompliance with the Observed – Expected threshold does not induce survival improvements, being near non-compliance leads those firms that do not shut down to improve their survival rates by a statistically significant 1.4 percentage points or a 16% reduction in the one year mortality rate (Table A16). Contrary to the overall results, we do not see a preemptive decrease in the transplant rate, overall or for deceased or live donor transplants. However, we do find a larger

preemptive effect for the rate at which patients are deemed too sick for transplant and a smaller (although still statistically significant) effect from actual noncompliance with this particular threshold. The effect on the transfer rate from noncompliance disappears in these regressions. The transfer effect in the overall regressions was not highly statistically significant, so part of the decline in significance could be due to the reduced sample size.

These results show that centers which do not shut down are able to preemptively improve one-year survival rates, exhibit a stronger response to the existence of enforcement (PostCMS variable), and show a slightly different pattern in the results with respect to removals for deceased donor and live donor transplants as well as removals for too sick for transplant or transfers. Unfortunately, we cannot run a similar analysis including only those centers that shutdown or exit due to the small number of centers in these categories. Attempts to include an interaction resulted in the term being dropped once center fixed effects are included.

Information in the data for tracking transfer patients across centers is noisy (patients sometimes disappear from the data set, presumably reappearing with a new patient identification code) and patient-level analysis did not find evidence that transfer patients were consistently sicker or healthier than other patients. However, information on patients wait-listed at multiple transplant centers provides some additional insight. Multi-listing is possible, although generally more expensive because most insurance companies will cover registration at only one transplant center. Multi-listing can be beneficial in that wait-list prioritization occurs first at the Donation Service Area-level (58 DSAs in the U.S.), so a patient listed in multiple DSAs will have “local” access to multiple DSA-level supplies of donor organs. In addition to the cost of multiple registrations, many patients are geographically limited in terms of where they can list because the patient must be able to arrive at the transplant center and be prepped for transplant within the period of time in which the donor kidney is viable – about 24 hours without a living blood supply. We use a delisting code for whether a multi-listed patient chooses to have a transplant at another center rather than at the center in question to determine whether arguably the wealthiest

and most mobile patients choose not to receive transplants at noncompliant transplant centers. Our results in Table A17 do not show any evidence of an increase in multi-listed patients who choose not to have a transplant at a given center.

We performed a variety of robustness checks on the inclusion of volume in order to determine whether it has an effect on our results (Table A18 through A21). Due to endogeneity concerns relating to selective referral (that better outcomes attract higher volume rather than vice versa), we report results with and without patient case mix. Without risk adjustment, what looks like a scale benefit could just be a reflection of larger centers treating healthier patients. In Tables A18 (no risk adjustment) and A19 (with risk adjustment), we included the raw number of patients and the natural log of the raw number of patients as independent variables for outcomes including the rate, count, and natural log of the count of one year survivors and still find no positive association between flagging and the number of one year post-transplant survivors. If anything, a negative relationship might exist. Using a three-way interaction between flagging, post-CMS, and measures of the volume or omitting center fixed effects (which can be correlated with scale if scale is fairly consistent over time) does not yield evidence of any survival benefit in Table A20. We also test an alternative specification (Table A21), which leads to selection issues, but avoids the endogeneity issue of selective referral (that better outcomes drive volume) arising from directly including volume in the regressions. (Here, too, we report results with and without risk adjustment.) If we restrict our sample to only those centers performing more than 35 transplants in a six-month period, we do find evidence of an increase in one year post-transplant survival with non-compliance in the post-CMS enforcement period. This cuts the number of flagged centers in the data by half with large centers having a slightly lower flagging rate (33% of centers with six-month volume of less than or equal to 35 and 27% of centers with six-month volume of greater than 35.) The counts are even lower in the post-CMS enforcement period with 43 of smaller centers flagged for noncompliance and only 29 of the larger centers. Because our functional sample (due to clustering at the transplant center level) is relatively small, especially for the post-CMS enforcement period, we are concerned about the validity and generalizability of our results for large centers. With the

inclusion of center fixed effects, we also do not have a strong argument for why within-center changes in scale alone would lead to improvements in one year post-transplant survival, especially because transplant centers are located in large not-for-profit hospitals, often are university-affiliated, and likely have similar resources available to them.

6. Conclusion

Both the public disclosure of information regarding the quality of healthcare outcomes and regulatory enforcement of performance standards should improve the quality of care provided due to increased competitive pressures from the market and the deterrent effect of regulatory penalties. With healthcare expenditures accounting for 17.2% of GDP in 2012,¹⁸ the quality of care provided is important both to ensure as healthy a population as possible and to avoid unnecessary and often expensive complications. Not surprisingly, insurers, including Medicare, use quality standards as one of the criteria for including hospitals or physicians within their network and to designate certain subgroups of hospitals as “Centers of Excellence.”

We study the impact of public quality reporting with and without enforcement. Most of the predictions from our simple supply and demand model (Figure 2) hold. Information alone has no impact in the presence of excess demand, the existence of enforcement has limited effects, but noncompliance or approaching noncompliance induces strategic behavior by transplant centers, which decrease the number of transplants and increase the number of patients deemed too sick for transplant or transferred to other centers. We do not, however, find any quality improvement despite evidence of changes in provider behavior.

The general lack of any survival improvement even with gaming behavior is perplexing. With one year post-transplant survival rates so high, it may be difficult for transplant centers to affect them. In fact, with one year survival rates currently around 92 percent, this measure approaches the threshold of 95 percent used by the Agency for Healthcare Research and Quality (AHRQ) for determining whether or not a quality measure should be retired.

¹⁸ <http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/downloads/highlights.pdf>. Accessed 06/1914.

Perhaps a shift toward the use of other post-transplant outcomes might be more beneficial, e.g., readmissions or rejection treatments, constraining the types of deaths to those most likely to be influenced by transplant center performance, or metrics based on the quality of follow-up.

Although survey evidence suggests centers attempt to improve quality when facing noncompliance (Schold et al. 2010), even highly sophisticated hospitals have struggled to determine what exactly caused them to become noncompliant, which means gaming behavior may not be nefarious but simply necessary. The head administrator of a major university hospital flagged for noncompliance disclosed that the center was never able to find the “smoking gun”, but that CMS was eventually satisfied (post-audit) with the increased quality controls they introduced, such as improving continuity of care across shifts, increasing the responsibilities and frequency with which their internal quality control committee meets and taking into account the input of all members of the transplant team, not just the surgeons. One year survival also stretches well past the patient’s discharge, meaning that improvements in patient follow-up may be the best way to improve outcomes by improving the quality of care received post-transplant and by responding more quickly to issues of patient noncompliance.

In general, survival rates have increased over time even absent CMS enforcement due to improvements in the efficacy of immunosuppression and the expansion of transplantation-specific training programs for surgeons, physicians, nurses, and administrators. Decreased emphasis of the role of individual surgeons relative to the transplant team also may have improved outcomes.¹⁹ A paper by Pisano, Bohmer, and Edmondson (2001) on learning-by-doing in minimally invasive coronary surgery provides anecdotal evidence that improving relationships and coordination at the team-level (e.g., doctors treating nurses with respect and better patient-management across shifts) leads to better patient outcomes. Given the

¹⁹ Increasingly, transplant centers are attempting to create general standards of care so that patients can be transplanted by any surgeon at the center when an organ becomes available, rather than relying on a specific surgeon-patient relationship, which requires both the surgeon and the patient to be permanently on call in the hope that an organ will become available.

team-based characteristics of transplantation, improved coordination among team members might improve outcomes.

The social welfare effects of these regulatory interventions are difficult to determine in such a complex environment, but we generally assess that information provision with or without enforcement will decrease welfare overall. The lack of any measurable change in the outcome variables from the introduction of information leads us to believe that such an intervention is not welfare-improving from a societal perspective.²⁰ Although patients' perception of the marginal benefit of a transplant at a given center may be more accurate, this does not affect outcomes overall, at compliant centers, or at noncompliant centers because the price still remains far below the marginal benefit perceived by most patients. It is unclear that demand shifts at all, and even if it did, it is uncertain whether it would shift in or out in aggregate. If anything, an increase in registrations in the information only period suggests a possible increase in demand, likely due to either the increasing incidence of ESRD or an increase in the perceived benefit of transplantation.

Enforcement might be welfare-improving if healthier patients and higher quality organs lead to more life years generated, even if fewer patients are transplanted. Our panel is too short to assess outcomes accurately post-enforcement for much more than a year or two, but this possibility could exist. However, enforcement also increases the marginal cost of performing transplants, and therefore, leads to a decrease in the supply of transplants or a decrease in equity due to the removal of patients now deemed too sick, who would have received a transplant prior to the introduction of the CMS enforcement. Based on the National Organ Transplant Act, we, as a society, have decided that we value equity in access to the cadaveric donor supply, so any increase in life years produced in transplantation must be weighed against welfare costs from decreased equity and access to transplantation. The kidney transplant community recently introduced a new kidney allocation algorithm after

²⁰ We abstract away from the obvious costs of reporting and data submission by hospitals, of data publication by the Scientific Registry of Transplant Recipients, of CMS oversight and enforcement efforts, and of patient and physician time spent researching and comparing center outcomes, which would further reduce any welfare benefits.

years of debate. “Life Years from Transplant” (LYFT) was a proposed metric, but was discarded as too discriminatory, especially due to the high correlation with wealth, and correspondingly, access to high quality healthcare. Preserving equity in the allocation of cadaveric donors may also help increase the organ supply. Surveys indicate that some members of minority populations choose to not donate because of the perception that the organs will just go to “rich white people” (Boulware et al., 2002). The LYFT system would likely only exacerbate such perceptions.

Non-compliance, relative to information only or simply the existence of enforcement, significantly increases the marginal cost of performing a transplant and thus leads to an even stronger response in terms of productive and allocative inefficiencies and decreased access to transplantation without a significant associated increase in the regulatory survival measure.

Any welfare improvements from additional LYFT must be weighed against potential cost increases from reduced access to transplantation, even apart from equity outcomes. Any reduction in the number of transplants carries with it high costs for both payers and patients. Matas and Schnitzler (2003) estimate that a kidney transplant saves more than \$90,000 per patient relative to maintenance dialysis²¹ and the Quality Adjusted Life Years (QALYs) associated with dialysis are significantly lower than those associated with transplantation (Wolfe et al., 1999). This does not include the costs associated with the decreased probability of employment for individuals on dialysis and the care provision required for an individual to remain on maintenance dialysis. (Most patients receive dialysis three times a week with each session lasting about four hours.²²) Our own preliminary work, a survey-based paper (Schold, 2010), and editorials in major transplant journals (Abecassis et al., 2009 and Roberts, 2012) also all point to potential dynamic inefficiencies resulting

²¹ Medicare gets an even bigger benefit because it covers patients only for three years post-transplant but covers dialysis patients for life or until the end of dialysis.

²² <http://www.mayoclinic.org/tests-procedures/hemodialysis/basics/what-you-can-expect/prc-20015015> Accessed 05/01/2015.

from reduced research activity and decreased adoption of new patient protocols as transplant centers attempt to reduce the overall risk of the transplants they perform.

Given the increased marginal cost to transplant centers of performing transplants, *ceteris paribus*, the higher cost of dialysis relative to transplantation, the reduction in the rate of transplantation/access to transplantation, and the societal value placed on equity in the distribution of the cadaveric donor supply, we believe that it is highly unlikely that either of the two types of regulatory interventions (information and enforcement) are welfare-improving and we may simply be trading off information-based market failures for “government failures.” The fact that these policies are not welfare-improving is likely not the result of a lack of information, but rather of government intervention, including the prohibition of organ sales. In the presence of a donation-only organ supply, removing the DRG-generated price ceiling and/or allowing for lower quality transplants may be the only effective way for the market to reach an equilibrium in which the demand for transplants equals the supply of transplants, although likely with reductions in equity. In the U.S. (and many other countries), we, as a society, have consistently expressed a preference for a donation-based organ supply making a market for kidneys in the U.S., the alternative solution to the current disparity between supply and demand, politically infeasible for the foreseeable future.

In summary, although intended to alleviate market failures associated with incomplete information, we find that the quality standards used in kidney transplantation do not generate a meaningful consumer response, but rather induce gaming behavior without any increase in quality. As a result, quality standards are likely to be social welfare-decreasing when, as in kidney transplantation, improvements in measured quality are difficult and prices and quality are not determined by the market.

References

Abecassis, M.M., R. Burke, G.B. Klintmalm, A.J. Matas, R.M. Merion, D. Millman, K. Olthoff, J.P. Roberts. 2009. American Society of Transplant Surgeons Transplant Center Outcomes

Requirements – A Threat to Innovation. *American Journal of Transplantation*, No. 9, pp.1279-1286.

Akerlof, George. 1970. The Market for “Lemons”: Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, Vol., 84, No. 3, pp.488-500.

Analysis of Changes to Transplantation Practices in the Period Surrounding the Initiation of Medicare Conditions of Participation for Organ Transplant Programs: Final Report. 2011. The University of Michigan Kidney Epidemiology and Cost Center: Ann Arbor, Michigan.

Bevan, Gwyn and Christopher Hood. 2006. What’s Measured Is What Matters: Targets and Gaming in the English Public Health Care System. *Public Administration*, Vol. 84, No. 3, pp.517-538.

Boulware, L.E., L.E. Ratner, J.A. Sosa, L.A. Cooper, R.A LaVeist, and N.R. Powe. 2002. Determinants of Willingness to Donate Living Related and Cadaveric Organs: Identifying Opportunities for Intervention. *Transplantation*, Vol. 73, pp.1683-1691.

Bundorf, M.K., N. Chun, G.S. Goda, D.P. Kessler. 2009. Do Markets Respond to Quality Information? The Case of Fertility Clinics. *Journal of Health Economics*, Vol. 28, pp.718-727.

Centers for Medicare and Medicaid Services. 2007. Medicare Program; Hospital Conditions for Participation: Requirements for Approval and Re-Approval of Transplant Centers to Perform Transplants; Final Rule. Title 42 Code of Federal Regulations, Parts 405, 482, 488, and 498. 2007. Federal Register, Vol.72, No. 61, pp.15198-15280.

Cox, James C.; Vjollca Sadiraj, Kurt E. Schnier, Nicole Turgeon. 2012. Loss Aversion and Organ Utilization. Working Paper.

Cutler, David M.; Robert S. Huckman, Mary Beth Landrum. 2004. The Role of Information in Medical Markets: An Analysis of Publicly Reported Outcomes in Cardiac Surgery. Working Paper. 2004.

Dranove, Daniel; Mark McClellan, and Mark Satterthwaite. 2003. Is More Information Better? The Effects of “Report Cards” on Health Care Providers. *Journal of Political Economy*, 111:555-588.

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*, Vol. 27, pp.1201-1207.

Eijkenaar, Frank. 2012. Pay for Performance in Health Care: An International Overview of Initiatives. *Medical Care Research and Review*, Vol. 69, No. 3, pp. 251-276.

Epstein, Andrew. 2006. "Do Cardiac Surgery Report Cards Reduce Mortality? Assessing the Evidence." *Medical Care Research and Review*, 63(4): 403-426.

Fisher, E.S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F.L. Lucas, E.L. Pinder. 2003. The Implications of Regional Variations in Medicare Spending, Part 1: The Content, Quality and Accessibility of Care. *Annals of Internal Medicine*, Vol. 138, No. 4, pp.273-287.

Heckman, James J., Carolyn J. Heinrich, Pascal Courty, Gerald Marschke, and Jeffrey Smith, Editors. 2011. The Performance of Performance Standards. Kalamazoo, Michigan: W.E. Upjohn Institute for Employment Research.

Kolstad, J. 2013. Information and Quality When Motivation Is Intrinsic: Evidence from Surgeon Report Cards. *American Economic Review*, Vol. 103, No. 7, pp. 2875-2910.

Hamilton, T.E. (2009) Accountability in Health Care – Transplant Community Offers Leadership. *American Journal of Transplantation*, No. 9, pp.1287-1293.

Howard, David H., and Bruce Kaplan. 2006. Do Report Cards Influence Hospital Choice? The Case of Kidney Transplantation. *Inquiry*, Vol. 43, No. 2, pp.150-159.

Matas, A. and M. Schnitzler. 2003. Payment for Living Donor (Vendor) Kidneys: A Cost-Effectiveness Analysis. *American Journal of Transplantation*, No. 4, pp.216-221.

Medicare's Coverage of Dialysis and Kidney Transplant Benefits: Getting Started. 2014. CMS Product No. 11360. <http://www.medicare.gov/Pubs/pdf/11360.pdf>. Accessed 06/18/14.

Mukamel, D.B., D.L. Weimer, J. Zwanziger, and A. Mushlin. 2002. Quality of Cardiac Surgeons and Managed Care Contracting Practices. *Health Services Research*, Vol. 37, No. 5, pp.1129-1144.

Nolte, Ellen, Martin Roland, Cheryl Damberg, Soeren Mattke, Mirella Cacace, Simo Goshev, Laura Brereton, Annalijn Conklin, Liisa Hiatt, Denise Quigley, and Susan Lovejoy. 2011. Informing the Development of a Resource Allocation Framework in the German Healthcare System. *Technical Report for the National Association of Statutory Health Insurance Physicians (Kassenärztliche Bundesvereinigung) – RAND Europe, RAND Corporation*, pp. 1-85.

Pisano, G., R.M.J. Bohmer, A.C. Edmondson. 2001. Organizational Differences in Rates of Learning: Evidence from the Adoption of Minimally Invasive Cardiac Surgery. *Management Science*, Vol. 47, No. 6, pp.752-768.

Roberts, John Paul. 2012. Impact of Outcomes Monitoring on Innovation and Risk in Liver Transplantation. *Liver Transplantation*, 18: S59-S63.

Roland, Martin, and Stephen Campbell. 2014. Successes and Failures of Pay for Performance in the United Kingdom. *The New England Journal of Medicine*, Vol. 370, No. 20, pp.1944-1949.

Schneider, Eric C., and Arnold M. Epstein. 1996. Influence of Cardiac-Surgery Performance Reports on Referral Practices and Access to Care: A Survey of Cardiovascular Specialists. *New England Journal of Medicine*, Vol. 335, No. 4, pp. 251-256.

Schnier, Kurt; C. McIntyre, R. Ruhil, V. Sadiraja, J.C. Cox, T.C. Pearson, A.D. Kirk, and N.A. Turgeon. 2013. The Impact of CMS CoP on Kidney Transplant Waiting Times. Working Paper.

Schold, Jesse D., Charlotte J. Arrington, Greg Levine (2010). Significant alterations in reported clinical practice associated with increased oversight of organ transplant center performance. *Progress in Transplantation*, Vol. 20, No.3, pp.279-287.

Steinbrook, Robert. 2006. "Public Report Cards -- Cardiac Surgery and Beyond." *New England Journal of Medicine*, 355(18): 1847-1849.

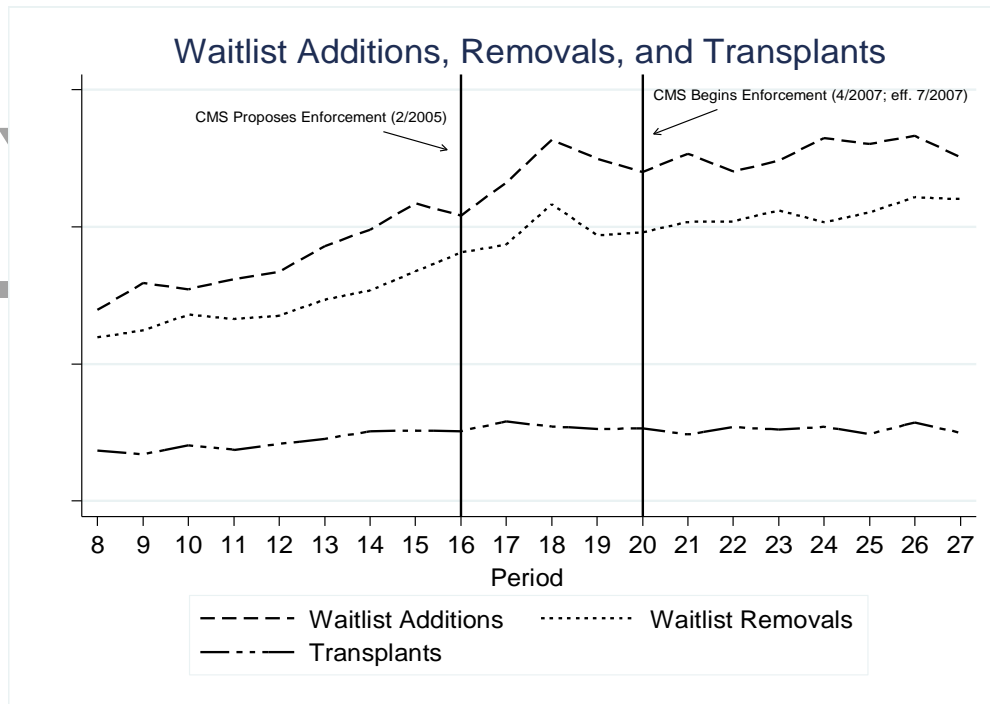
Tanriover B1, Mohan S, Cohen DJ, Radhakrishnan J, Nickolas TL, Stone PW, Tsapepas DS, Crew RJ, Dube GK, Sandoval PR, Samstein B, Dogan E, Gaston RS, Tanriover JN, Ratner LE, Hardy MA. 2014. Kidney at Higher Risk of Discard: Expanding the Role of Dual Kidney Transplantation. *American Journal of Transplantation*, Vol. 14, No. 2, pp.404-415.

Wang, Justin, Jason Hockenberry, Shin-Yi Chou, Muzhe Yang. 2011. Do Bad Report Cards Have Consequences? Impacts of Publicly Reported Provider Quality Information on the CABG market in Pennsylvania. *Journal of Health Economics*, Vol. 30, pp.392-407.

Werner, R.M., E.C. Norton, R.T. Konetzka, and D. Polsky. 2012. Do Consumers Respond to Publicly Reported Information? Evidence from Nursing Homes. *Journal of Health Economics*, Vol. 31, pp.50-61.

Wolfe, R.A., V.B. Ashby, E.L. Milford, A.O. Ojo, R.E. Ettenger, L.Y.C. Agodoa, P.J. Held, and F.K. Port. 1999. Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *New England Journal of Medicine*, Vol. 341, pp. 1725–1730.

Figure 1: Supply and Demand for Transplants by Period



ar

Figure 2: Market for Kidney Transplants

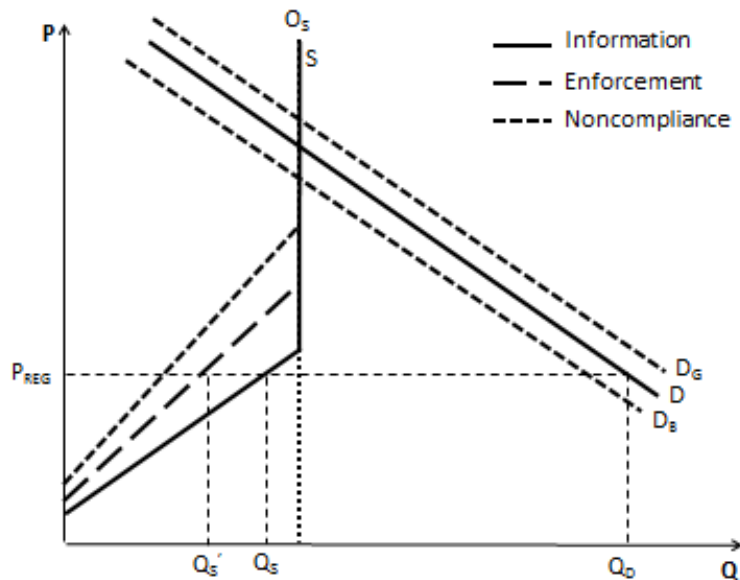


Figure 3: Data Structure

t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10	t=11	t=12	t=13	t=14	t=15
Cohort 1							R							
	Cohort 2						R							
		Cohort 3						R						
			Cohort 4						R					
				Cohort 5						R				

Figure 4: Flagging Incidence

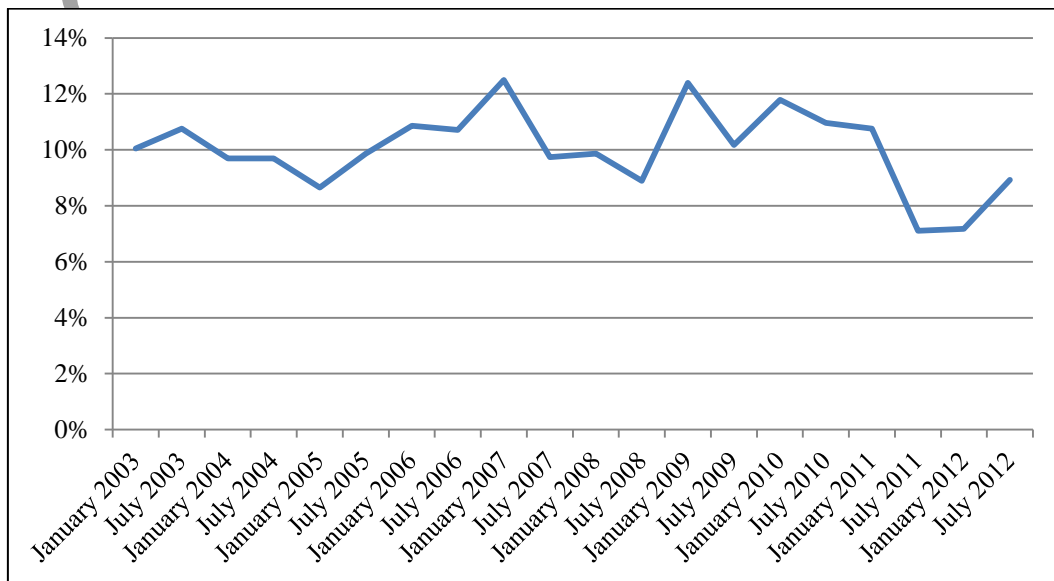
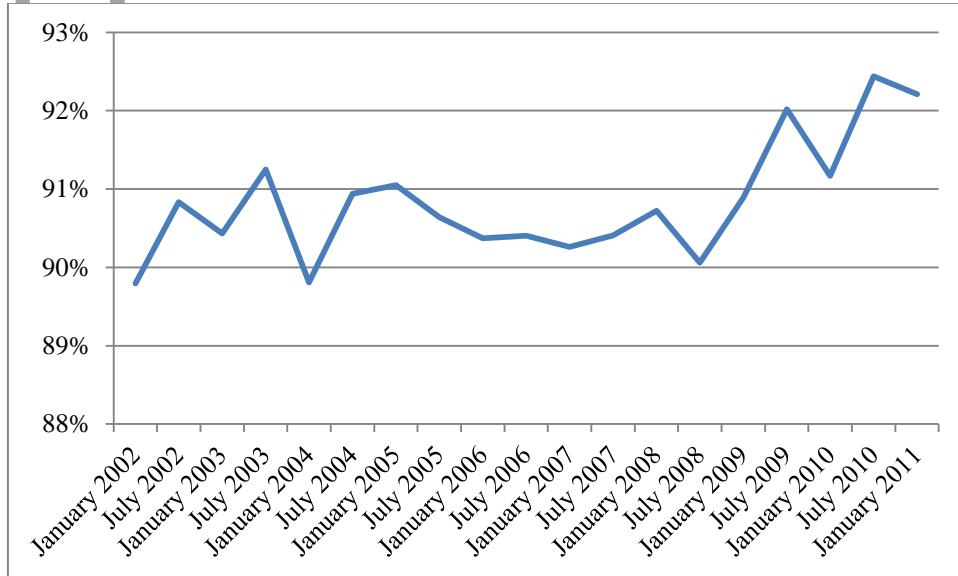


Figure 5: One Year Post-Transplant Survival Rate



Author Manuscript

Table 1: Flagging and Market Participation

	Entry	Shutdown	Exit	Re-Entry
Total Centers	15	34	23	11
# Flagged the period before	.	7	5	1
# Flagged within the prior two periods	.	12	10	1
# Flagged within the prior three periods	.	15	13	2
Average # of flags	.	1.4	2	0.2
Minimum # of flags	.	0	0	0
Maximum # of flags	.	11	11	2

Table 2: Descriptive Statistics for Outcome Variables

Variable	Observations	Mean	Standard Deviation*
Total transplants	4311	34	29
One month survival rate	4311	98%	6%
One year survival rate	4311	91%	12%
Three year survival rate	4311	68%	29%
Waitlist Removals	4482	62	69
Removal rate for transplants	4482	57%	20%
Removal rate for deceased donors	4482	41%	19%
Removal rate for live donors	4482	15%	12%
Removal rate for too sick for transplant	4482	5%	7%
Rate of transfers to other centers	4482	13%	16%

*The percentages are percentage point differences from the mean, not percentages of the mean.

Notes: The data include all observations for which flagging information is available.

Table 3: One Year Post-Transplant Survival Rate

	t	$t+1$	$t+2$	$t+3$
Flagged in period t	0.010 (0.012)	0.001 (0.011)	0.005 (0.011)	0.010 (0.011)
Post-CMS oversight	0.009 (0.008)	0.008 (0.008)	0.004 (0.008)	0.006 (0.009)
Flag * Post-CMS	-0.014 (0.019)	-0.004 (0.015)	0.008 (0.015)	0.007 (0.018)
Observations	4,311	4,076	3,836	3,605
R-squared	0.247	0.243	0.246	0.268

Notes: Regressions include center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4: Transplant Rate

	<i>t</i>	<i>t+1</i>	<i>t+2</i>	<i>t+3</i>
Flag for cohort ending at <i>t</i>	0.024 (0.015)	0.016 (0.016)	0.012 (0.018)	0.024 (0.018)
Post-CMS oversight	-0.015 (0.010)	-0.022** (0.010)	-0.019* (0.011)	-0.018* (0.011)
Flag * Post-CMS	-0.074*** (0.026)	-0.064** (0.025)	-0.056** (0.027)	-0.062** (0.025)
Observations	4,482	4,236	3,989	3,750
R-squared	0.437	0.437	0.458	0.462

Notes: Regressions include center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 5: Deceased Donor Transplant Rate

	<i>t</i>	<i>t+1</i>	<i>t+2</i>	<i>t+3</i>
Flag for cohort ending at <i>t</i>	0.020 (0.014)	0.022 (0.015)	0.015 (0.017)	0.022 (0.017)
Post-CMS oversight	-0.017* (0.009)	-0.023** (0.009)	-0.020** (0.009)	-0.017* (0.009)
Flag * Post-CMS	-0.049** (0.024)	-0.043* (0.024)	-0.035 (0.024)	-0.041* (0.023)
Observations	4,482	4,236	3,989	3,750
R-squared	0.433	0.426	0.447	0.441

Notes: Regressions include center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 6: Live Donor Transplant Rate

	<i>t</i>	<i>t+1</i>	<i>t+2</i>	<i>t+3</i>
Flag for cohort ending at <i>t</i>	0.003 (0.006)	-0.006 (0.006)	-0.003 (0.006)	0.002 (0.007)
Post-CMS oversight	0.002 (0.005)	0.001 (0.006)	0.001 (0.006)	-0.001 (0.006)
Flag * Post-CMS	-0.025** (0.011)	-0.022** (0.010)	-0.020** (0.009)	-0.020** (0.009)
Observations	4,482	4,236	3,989	3,750
R-squared	0.373	0.379	0.390	0.401

Notes: Regressions include center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 7: Too Sick for Transplant Removal Rate

	<i>t</i>	<i>t+1</i>	<i>t+2</i>	<i>t+3</i>
Flag for cohort ending at <i>t</i>	-0.001 (0.005)	-0.005 (0.005)	-0.005 (0.006)	-0.009* (0.005)
Post-CMS oversight	0.006 (0.004)	0.005 (0.004)	0.004 (0.004)	-0.000 (0.004)
Flag * Post-CMS	0.018** (0.007)	0.025*** (0.008)	0.017* (0.009)	0.021** (0.009)
Observations	4,482	4,236	3,989	3,750
R-squared	0.269	0.275	0.278	0.288

Notes: Regressions include center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 8: Transferred to Another Center Removal Rate

	<i>t</i>	<i>t+1</i>	<i>t+2</i>	<i>t+3</i>
Flag for cohort ending at <i>t</i>	-0.020** (0.009)	-0.003 (0.011)	-0.012 (0.011)	-0.016 (0.010)
Post-CMS oversight	-0.007 (0.010)	-0.002 (0.010)	-0.004 (0.010)	-0.004 (0.011)
Flag * Post-CMS	0.031* (0.017)	0.014 (0.018)	0.028* (0.014)	0.029* (0.017)
Observations	4,482	4,236	3,989	3,750
R-squared	0.374	0.388	0.385	0.380

Notes: Regressions include center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 9: Descriptive Statistics for Categories of Individual Measures

Variable	Observations*	Mean	Standard Deviation**	Number of Centers**
Observed - Expected	4497	0.01	3.09	266
Observed - Expected ≤ 0	4497	0.57	0.50	257
$0 < \text{Observed} - \text{Expected} \leq 3$	4497	0.31	0.46	220
Observed - Expected > 3	4497	0.12	0.33	115
Observed - Expected > 3 (but still compliant)	4046	0.02	0.15	33

*This only includes observations starting in period 8, when first report was issued.

**This is the number of centers that fall into a given category for at least one period.

Table 10: Effects of Magnitude of Difference between Expected and Observed Deaths

	One Year Post Transplan t Survival Rate	Transplan t Rate	Deceased Donor Transplan t Rate	Live Donor Transplan t Rate	Too Sick for Transplan t Rate	Transfer Rate
0 < Observed - Expected ≤ 3	0.004 (0.006)	0.013 (0.009)	0.008 (0.009)	0.005 (0.006)	-0.000 (0.004)	-0.017** (0.008)
Observed - Expected > 3	0.012 (0.010)	0.024 (0.015)	0.022 (0.014)	0.002 (0.006)	-0.001 (0.005)	-0.023** (0.009)
Post-CMS	0.007 (0.007)	0.001 (0.011)	-0.008 (0.010)	0.008 (0.006)	0.005 (0.004)	-0.021** (0.010)
0 < O-E ≤ 3*PostCMS	0.002 (0.010)	-0.042*** (0.015)	-0.024* (0.014)	-0.017** (0.008)	0.002 (0.007)	0.037** (0.015)
O-E > 3*PostCMS	-0.011 (0.017)	-0.079*** (0.025)	-0.052** (0.024)	-0.027** (0.011)	0.019** (0.007)	0.039** (0.016)
Observations	4,273	4,482	4,482	4,482	4,482	4,482
R-squared	0.257	0.438	0.433	0.373	0.270	0.376

Notes: Regressions include center fixed effects and a six month period level time trend. The omitted "Observed - Expected" category is Observed - Expected ≤ 0. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A1: Market Participation Decisions by Period

Period	Entry	Shutdown	Exit	Re-Entry
8	0	2	2	0
9	2	0	0	0
10	1	2	1	1
11	0	1	1	0
12	0	1	1	0
13	1	1	0	1
14	1	3	3	1
15	1	2	1	1
16	1	2	2	0
17	0	0	0	1
18	0	4	3	1
19	4	1	1	0
20	2	5	4	1
21	1	1	1	0
22	0	3	1	1
23	1	3	1	0
24	0	0	0	2
25	0	2	0	0
26	0	1	1	0
27	0	0	0	1
Total	15	34	23	11

Table A2: Cumulative Flags and Period of Exit for Each Exiting Center

Exit Period	Cumulative Flags
8	0
8	1
10	0
11	0
12	1
14	0
14	3
14	2
15	0
16	1
16	0
18	0
18	0
18	0
19	0
20	2
20	0
20	1
20	2
21	11
22	2
23	10
26	9

Author Manuscript

Table A3: Total Flags

Max # of Flags	# of Centers
0	136
1	25
2	15
3	16
4	6
5	7
6	6
7	5
8	7
9	4
10	2
11	3
12	1
13	3
14	1

Table A4: Consecutive Flags

# of Flags in a Row*	Count
1	73
2	39
3	15
4	7
5	13
6	7
7	8
9	4
10	1
13	1

*These are exclusive categories

Table A5: One Year Post-Transplant Survival with Risk Adjustment

	<i>t</i>	<i>t+1</i>	<i>t+2</i>	<i>t+3</i>
Flagged in period <i>t</i>	0.009 (0.011)	-0.000 (0.009)	0.003 (0.010)	0.001 (0.009)
PostCMS	-0.001 (0.007)	-0.001 (0.007)	-0.002 (0.007)	-0.001 (0.007)
Flag * PostCMS	-0.005 (0.016)	-0.000 (0.014)	0.011 (0.014)	0.018 (0.017)
Observations	4,073	3,859	3,639	3,425
R-squared	0.380	0.392	0.368	0.386

Notes: Regressions include center fixed effects, a six month period level time trend, and casemix based on SRTR methodology. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A6: Negative Binomial Regressions for Total Transplants

	<i>t</i>	<i>t+1</i>	<i>t+2</i>	<i>t+3</i>
Flagged in period <i>t</i>	1.176 (1.260)	1.468 (1.344)	1.847 (1.561)	1.617 (1.476)
Post-CMS oversight	5.769*** (0.837)	4.740*** (0.845)	4.476*** (0.806)	3.851*** (0.804)
Flag * Post-CMS	-9.107*** (2.567)	-6.182** (2.911)	-7.345** (3.032)	-9.256*** (2.761)
Observations	3,888	3,653	3,415	3,186

Notes: Reported coefficients are average marginal effects. Regressions include center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A7: Negative Binomial Regressions for One Year Post-Transplant Survivors

	<i>t</i>	<i>t+1</i>	<i>t+2</i>	<i>t+3</i>
Flagged in period <i>t</i>	0.634 (0.760)	0.467 (0.821)	0.779 (0.961)	0.909 (0.865)
Post-CMS oversight	0.694 (0.508)	0.639 (0.517)	0.629 (0.495)	0.898* (0.509)
Flag * Post-CMS	-3.793*** (1.408)	-4.107** (1.628)	-2.594 (1.613)	-5.257*** (1.598)
Observations	3,888	3,653	3,415	3,186

Notes: Reported coefficients are average marginal effects. Regressions include total transplants each period, center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A8: One Month Post-Transplant Survival Rate

	<i>t</i>	<i>t+1</i>	<i>t+2</i>	<i>t+3</i>
Flagged in period <i>t</i>	0.000 (0.005)	0.001 (0.005)	0.005 (0.007)	0.003 (0.006)
Post-CMS oversight	0.000 (0.005)	0.003 (0.004)	0.001 (0.005)	0.001 (0.005)
Flag * Post-CMS	-0.006 (0.010)	-0.015 (0.012)	-0.002 (0.008)	-0.006 (0.012)
Observations	4,311	4,076	3,836	3,605
R-squared	0.129	0.131	0.134	0.140

Notes: Regressions include center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A9: Three Year Post-Transplant Survival

	<i>t</i>	<i>t+1</i>	<i>t+2</i>	<i>t+3</i>
Flagged in period <i>t</i>	0.021** (0.011)	0.013 (0.010)	0.012 (0.012)	0.007 (0.012)
Post-CMS oversight	0.001 (0.012)	0.006 (0.012)	0.004 (0.012)	0.003 (0.012)
Flag * Post-CMS	0.015 (0.026)	0.004 (0.028)	-0.005 (0.032)	-0.009 (0.031)
Observations	3,459	3,226	2,991	2,766
R-squared	0.455	0.472	0.481	0.512

Notes: Regressions include center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A10: Flagging Incidence by Type of Noncompliance

	Mean Percent of Centers Flagged in a Period	Number of Centers Ever Flagged	Percent of Centers Ever Flagged
Flag - Patient Deaths	10%	102	35%
Flag - Graft Failures	11%	116	40%
Flag - Either	15%	134	46%
Flag - Both	5%	73	25%

Table A11: Effects of Alternative Flagging Measures

Flagged for:	Patient Deaths	Graft Failures	Both
One Year Post-Transplant Survival Rate			
Flag * PostCMS	-0.011 (0.019)	-0.024 (0.025)	-0.042 (0.036)
Transplant Rate			
Flag * PostCMS	-0.074*** (0.026)	-0.078*** (0.027)	-0.093** (0.037)
Deceased Donor Transplant Rate			
Flag * PostCMS	-0.049** (0.024)	-0.064** (0.027)	-0.069** (0.034)
Live Donor Transplant Rate			
Flag * PostCMS	-0.025** (0.011)	-0.014 (0.011)	-0.023* (0.013)
Too Sick for Transplant Rate			
Flag * PostCMS	0.018** (0.007)	0.019** (0.008)	0.015 (0.011)
Transfer Rate			
Flag * PostCMS	0.031* (0.017)	0.028** (0.013)	0.047* (0.025)

Notes: Outcomes are in the period at the beginning of which the SRTR report was issued (time = t). Regressions include center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A12: The Effect of First-Time Noncompliance

Outcome	One Year Post Transplant Survival Rate	Transplant Rate	Deceased Donor Transplant Rate	Live Donor Transplant Rate	Too Sick for Transplant Rate	Transfer Rate
Flagged in period t	0.007 (0.016)	0.004 (0.016)	0.005 (0.015)	-0.001 (0.010)	0.009 (0.007)	-0.022 (0.015)
Post-CMS oversight	0.014 (0.009)	-0.011 (0.011)	-0.013 (0.010)	0.002 (0.006)	0.006 (0.004)	-0.011 (0.011)
Flag * Post-CMS	0.006 (0.024)	-0.043* (0.026)	-0.034 (0.022)	-0.009 (0.016)	0.020 (0.014)	0.014 (0.021)
Observations	3,150	3,591	3,591	3,591	3,591	3,591
R-squared	0.255	0.457	0.458	0.350	0.287	0.384

Notes: Outcomes are in the period at the beginning of which the SRTR report was issued. Regressions include center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A13: Omitting Nonconsecutive Flags

	One Year Post Transplant Survival Rate	Transplant Rate	Deceased Donor Transplant Rate	Live Donor Transplant Rate	Too Sick for Transplant Rate	Transfer Rate
Flagged in period t	0.012 (0.012)	0.015 (0.014)	0.012 (0.014)	0.003 (0.006)	-0.003 (0.005)	-0.008 (0.006)
Post-CMS oversight	0.007 (0.008)	-0.013 (0.010)	-0.016* (0.009)	0.003 (0.006)	0.008* (0.004)	-0.016* (0.009)
Flag * Post-CMS	-0.016 (0.020)	-0.055** (0.023)	-0.028 (0.021)	-0.027** (0.012)	0.018** (0.009)	0.014 (0.011)
Observations	3,847	3,962	3,962	3,962	3,962	3,962
R-squared	0.247	0.421	0.432	0.353	0.284	0.385

Notes: Regressions include center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A14: Omitting More Than Three Consecutive Flags

	One Year Post Transplant Survival Rate	Transplant Rate	Deceased Donor Transplant Rate	Live Donor Transplant Rate	Too Sick for Transplant Rate	Transfer Rate
Flagged in period t	0.008 (0.018)	0.004 (0.013)	0.007 (0.013)	-0.003 (0.010)	0.001 (0.009)	-0.014* (0.007)
Post-CMS oversight	0.009 (0.008)	-0.023** (0.009)	-0.025*** (0.009)	0.002 (0.006)	0.006 (0.004)	-0.002 (0.006)
Flag * Post-CMS	0.002 (0.030)	-0.051** (0.024)	-0.049** (0.024)	-0.002 (0.018)	0.012 (0.015)	0.027 (0.022)
Observations	3,184	3,183	3,183	3,183	3,183	3,183
R-squared	0.307	0.490	0.532	0.467	0.302	0.371

Notes: Regressions include center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A15: Omitting Centers That Shutdown

	One Year Post Transplant Survival Rate	Transplant Rate	Deceased Donor Transplant Rate	Live Donor Transplant Rate	Too Sick for Transplant Rate	Transfer Rate
Flagged in period t	0.002 (0.007)	0.011 (0.014)	0.008 (0.014)	0.002 (0.006)	0.003 (0.005)	-0.007 (0.006)
Post-CMS oversight	0.002 (0.004)	-0.017* (0.009)	-0.019** (0.008)	0.002 (0.005)	0.011*** (0.004)	-0.005 (0.004)
Flag * Post-CMS	-0.002 (0.015)	-0.045** (0.023)	-0.019 (0.021)	-0.026** (0.012)	0.011 (0.008)	0.004 (0.009)
Observations	3,713	3,713	3,713	3,713	3,713	3,713
R-squared	0.205	0.485	0.512	0.501	0.304	0.421

Notes: Regressions include center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A16: Preemptive Effects Omitting Centers That Shutdown

	One Year Post Transplan t Survival Rate	Transpla nt Rate	Decease d Donor Transpla nt Rate	Live Donor Transpla nt Rate	Too Sick for Transpla nt Rate	Transfer Rate
$0 < \text{Observed} - \text{Expected} \leq 3$	-0.003 (0.003)	-0.005 (0.007)	-0.005 (0.007)	-0.000 (0.005)	-0.003 (0.003)	0.001 (0.004)
Observed - Expected > 3 (noncompliant)	-0.001 (0.006)	0.006 (0.014)	0.007 (0.014)	-0.001 (0.006)	0.002 (0.005)	-0.006 (0.005)
Post-CMS	-0.003 (0.005)	-0.014 (0.010)	-0.018** (0.009)	0.005 (0.006)	0.007* (0.004)	-0.007 (0.005)
$0 < O-E \leq 3 * \text{PostCMS}$	0.014*** (0.005)	-0.008 (0.011)	-0.001 (0.011)	-0.008 (0.007)	0.011** (0.005)	-0.001 (0.005)
$O-E > 3 * \text{PostCMS}$	0.006 (0.012)	-0.042* (0.022)	-0.019 (0.020)	-0.023** (0.011)	0.014* (0.007)	0.003 (0.008)
Observations	3,713	3,713	3,713	3,713	3,713	3,713
R-squared	0.207	0.485	0.513	0.501	0.306	0.425

Notes: Regressions include center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A17: Removed for Transplant at Another Center (Multi-listings Only)

	<i>t</i>	<i>t+1</i>	<i>t+2</i>	<i>t+3</i>
Flagged in period <i>t</i>	-0.002 (0.004)	-0.002 (0.004)	-0.000 (0.004)	0.004 (0.005)
Post-CMS oversight	-0.001 (0.004)	-0.004 (0.004)	-0.005 (0.005)	-0.004 (0.005)
Flag * Post-CMS	-0.004 (0.007)	-0.006 (0.006)	-0.004 (0.006)	-0.006 (0.007)
Observations	4,482	4,236	3,989	3,750
R-squared	0.375	0.375	0.385	0.412

Notes: Regressions include center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A18: One Year Post-Transplant Survival Controlling for Volume

Outcome	One Year Survival Rate	Number of One Year Survivors	Ln(Number of One Year Survivors)	One Year Survival Rate	Number of One Year Survivors	Ln(Number of One Year Survivors)
Flagged in period <i>t</i>	0.009 (0.012)	0.114 (0.193)	0.027 (0.019)	0.010 (0.011)	0.544 (0.640)	0.027* (0.015)
PostCMS oversight	0.009 (0.008)	0.046 (0.131)	0.005 (0.021)	0.008 (0.008)	-1.916*** (0.515)	-0.069*** (0.014)
Flag*PostCMS	-0.014 (0.019)	0.167 (0.340)	-0.044 (0.037)	-0.005 (0.018)	-3.689** (1.697)	-0.094** (0.040)
Number of transplants	-0.000 (0.000)	0.902*** (0.006)	0.024*** (0.002)			
Ln(Number of transplants)				0.009 (0.006)	12.310*** (0.745)	0.668*** (0.018)
Observations	4,291	4,291	4,258	3,888	3,888	3,857
R-squared	0.251	0.995	0.935	0.251	0.925	0.952

Notes: Regressions include center fixed effects and a six month period level time trend. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A19: One Year Post-Transplant Survival Controlling for Volume and Case Mix

Outcome	One Year Survival Rate	Number of One Year Survivors	Ln(Number of One Year Survivors)	One Year Survival Rate	Number of One Year Survivors	Ln(Number of One Year Survivors)
Flagged in period t	0.009 (0.011)	0.152 (0.193)	0.017 (0.018)	0.009 (0.011)	0.391 (0.622)	0.016 (0.015)
PostCMS oversight	-0.001 (0.007)	0.024 (0.143)	-0.011 (0.019)	-0.002 (0.006)	-0.888* (0.492)	-0.044*** (0.013)
Flag*PostCMS	-0.005 (0.016)	0.133 (0.342)	-0.031 (0.036)	0.003 (0.016)	-3.760** (1.606)	-0.082** (0.040)
Number of transplants	0.000 (0.000)	0.905*** (0.006)	0.024*** (0.001)			
Ln(Number of transplants)				0.012** (0.005)	13.163** *	0.645*** (0.018)
Observations	4,073	4,073	4,050	3,676	3,676	3,654
R-squared	0.380	0.995	0.929	0.391	0.926	0.940

Notes: Regressions include center fixed effects, a six month period level time trend, and casemix controls based on the SRTR risk adjustment methodology. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A20: One Year Post-Transplant Survival with Volume Interactions and Omitting Center Fixed Effects

Controls	Interactions		No Center FE	
	Number of transplants	Ln(Number of transplants)	No Risk Adj.	Risk Adj.
Flagged in period t	0.023 (0.024)	-5.137* (3.036)	-0.012 (0.008)	-0.012 (0.008)
PostCMS oversight	0.007 (0.010)	0.086 (1.548)	0.014* (0.008)	0.001 (0.007)
Flag * PostCMS	-0.049 (0.040)	9.160 (5.823)	-0.013 (0.014)	0.001 (0.013)
Number of transplants	-0.000 (0.000)			
Flag * Number of transplants	-0.000 (0.000)			
PostCMS * Number of transplants	0.000 (0.000)			
Flag * PostCMS * Number of transplants	0.001 (0.001)			
Ln(Number of transplants)		12.739*** (0.851)		
Flag * Ln(Number of transplants)		1.750* (0.999)		
PostCMS * Ln(Number of transplants)		-0.641 (0.479)		
Flag * PostCMS * Ln(Number of transplants)		-3.828* (2.132)		
Observations	4,291	3,888	4,291	4,073
R-squared	0.252	0.926	0.005	0.076

Notes: Regressions include center fixed effects and a six month period level time trend; the regression in the last column also includes casemix based on SRTTR methodology. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A21: One Year Post-Transplant Survival with Subsamples of Centers by Volume

	(1)	(2)	(3)	(4)	(5)	(6)
Transplants per period > N	N=35	N=30	N=25	N=35	N=30	N=25
Flag in period t	-0.000 (0.005)	0.000 (0.005)	0.000 (0.006)	-0.003 (0.005)	-0.002 (0.005)	-0.001 (0.005)
PostCMS	-0.002 (0.004)	-0.002 (0.004)	-0.003 (0.004)	-0.003 (0.005)	-0.002 (0.004)	-0.004 (0.005)
Flag * PostCMS	0.018** (0.008)	0.013 (0.008)	0.008 (0.010)	0.022*** (0.008)	0.017** (0.008)	0.012 (0.010)
Observations	1,621	1,922	2,269	1,596	1,893	2,237
R-squared	0.467	0.452	0.427	0.506	0.493	0.458

Notes: Regressions include center fixed effects and a six month period level time trend. Regressions in Columns [4]-[6] include casemix based on SRTR methodology. Standard errors are clustered at the center level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$