Do transplant centers change strategies after a poor performance?*

Han Ng†

this version: April 2024

Click here for the latest draft.

Abstract

Medicare's conditions of participation (CoP) is a policy that requires kidney transplant centers' numbers of graft failure or patient death 1 year after transplant to fall below a cutoff. Centers that repeatedly exceed the cutoff are flagged for poor performance and risk losing Medicare funding or certification. I use a sharp regression discontinuity design to study centers' response to being flagged for poor performance. Contrary to the existing literature, I find no evidence to suggest that flagged centers reduce (increase) the transplant of high (low) risk kidneys or waitlist younger, less obese or non-diabetic patients.

*I am very grateful for the guidance and encouragement from Paul Grieco, Mark Roberts and Karl Schurter. I thank Marc Henry and participants of Econ 555 in Penn State University for their valuable feedback. All errors are my own.

†Department of Economics, Pennsylvania State University, Kern Building, University Park, PA 16802, hln14@psu.edu.

1 Introduction

91814 patients were on the U.S. kidney waitlist in 2019 but only 16534 received a transplant and 8013 either died or became unsuitable for transplant while being in the waitlist[1]. Each transplant improves the expected quality and length of a transplanted patient's life (Wolfe et al., 2008). However, transplant centers have been accused of denying patietns treatment to meet federal standards [2]. These standards are part of Medicare's condition of particiation (CoP henceforth) that examines a transplant center's number of graph failure of patient death 1 year after transplant. Opponents argue that CoP incentivizes transplant centers to "game" the system by selectively waitlisting healthier patients or spending too much time waiting for the best kidneys (Schold, 2020).

My paper uses center-level administrative data to study the role of CoP in shaping center incentives in the U.S. deceased donor kidney transplant. Medicare enacted the CoP in May 2007. Under this policy, the penalty for underperformance can be severe. Centers that were flagged twice in 30 months are given a brief probationary period to improve or convince regulators that there were mitigating factors affecting their performance. Otherwise, the centers risk losing their certification or funding from Medicare.

As Medicare is the largest purchaser of solid organ transplantation , the prospect of Medicare withdrawal can weigh in on a center's transplant or waitlist strategy (Hamilton, 2013). In fact, a *New York Times* journalist interviewed the director of a kidney transplant program and he had this to say about Medicare's CoP [3]:

> "*...When you are looking at organs on the margins, if you have had a couple of bad outcomes recently, you say: 'Well why should I do this?'... You can always find a reason to turn organs down. It is this whole cascade that winds up with*

[1]Source: https://optn.transplant.hrsa.gov/data/view-data-reports/national-data/

[2]Source: https://www.statnews.com/2016/08/11/organ-transplant-federal-standards/

[3]Source:https://www.nytimes.com/2012/09/20/health/transplant-experts-blame-allocation-system-for-discarding-kidneys.html

people denied care or with reduced access to care."

This suggests that the performance concerns generated by Medicare's CoP indirectly force centers to ration health care under the guise of quality. This is potentially bad for patients. For example, non-waitlisted patients lose an opportunity to receive a new kidney. Secondly, patients on the waitlist might end up staying longer than necessary and become unsuitable for transplant (Held et al., 2016).

In this paper, I study the response of kidney transplant centers to being flagged by CoP for poor performance. In particular, I examine if flagged centers (i)reduce (increase) the transplant of high (low) risk kidneys; (ii) waitlist younger, less obese and non-diabetic patients. I use a sharp regression discontinuiuty (RD henceforth) design, relying on the fact that centers are only flagged if their numbers of graft failure or patient death exceed a cutoff. This allows me to obtain credible estimates of the average teatment effect (ATE henceforth) of being flagged at the cutoff. Furthermore, I implement McCrary (2008) density test and find no evidence of strategic behavior of the running variables near the cutoff. THis gives me confidence that my RD design is valid and satisfy the no manipulation assumption formalized in Imbens and Lemieux (2008).

My main findings suggest there is little evidence of flagged centers changing their transplant or waitlist strategy. This is in contrast to the results of Schold et al. (2013) and White et al. (2015). THese papers compare outcomes of flagged and non-flagged center without addressing endogeneity issues. For example, low-ability centers might not have the best facilities or surgeons, making them more susceptible to being flagged by CoP. THe difference in abilities also constrain the types of transplants centers attemot. The RD design overcomes these issues and uses exogenous variation that arise from discontinuity in CoP flagging rules. Randomization of flaggin status near the cutoff allows me to identify the ATE estimates. I also present suggestive evidence to support the nonparametric approach over the parametric approach in my setting.

My paper is related to two strands of literature. First, my paper constributes to existing works on the US deceased donor kidney transplant. Zhang (2010) and Agarwal et al. (2021) model patient decision to accept or reject a kidney as a dynamic optimization problem. Most directly related are Schold et al. (2013) and White et al. (2015). These papers document several stylized facts for deceased donor kidney transplant following CoP's introduction in 2007 and compare outcomes between flagged and non-flagged centers. Stith and Hirth (2016) uses a difference-in-differnces approach to analyze how the introduciong of CoP has affected patient survival. I contribute to this literature by introducing existing methods from the RD literature, see surveys by Imbens and Lemieux (2008) and Lee and Lemieux (2010), to exploit the discontinuity in CoP"s implementation and identify COP's effects on center behavior.

Secondly, there is a huge empirical literature looking at the effects of quality disclosure in the various industries within healthcare.Dranove et al. (2003),Kolstad (2013) looks at Coronary Artery Bypass Graft (CABG); Ramanarayanan (2011) studies dialysis centers; Bundorf et al. (2009) examines fertility clinics. Readers interested in this literature can refer to the survey by Dranove and Jin (2010). My paper on CoP in the U.S. deceased donor kidney transplant adds to this rich literature.

2 Background and Institutional Setting

This paper looks specifically at the U.S. deceased donor kidney transplant setting that forms approximately 70% of all kidney transplants in the U.S (AKF, 2008)[4]. I refer readers interested in the biology of kidney transplant and geography of the kidney market to the Appendix.

[4]The remaining 30% are either living donor kidney transplants or kidney exchange.

2.1 Registration at Transplant Centers

When a patient has kidney failure, his nephrologist refers him to a local transplant center. An evaluation will help determine if the patient is a suitable candidate for kidney transplant. The center's selection committee will review and decide if they want to accept the eligible candidate. Accepted patients will then be registered and placed on the waitlist [5].

2.2 Deceased Donor Kidney Allocation Process

Deceased donor kidneys are allocated through a centralized waitlist. When a deceased donor kidney becomes available, the centralized system identifies local patients that are biologically compatible and ranks them according to the amount of time spent on the waitlist (i.e. 1st patient spent the most time on the waitlist). Transplant centers are informed simultaneously of all their compatible patients and have 1 hour to decide on a provisional acceptance or immediate rejection[6] A compatible candidate can only receive the kidney if all candidates ranked before him decline the kidney. If the kidney is rejected by all local candidates, it will be offered to other patients in the U.S.

2.3 Conditions of Participation (CoP)

CoP is a list of performance criteria established by Medicare in May 2007. It is compulsory for transplant centers to submit reports on their transplant activities and outcomes bianually on every December and June. The Scientific Registry of Transplant Recipient (SRTR) collects these information and publishes reports on the second Tuesday of the immediate January

[5]Multiple listings are permitted but rarely adopted. This is because patients generally have to register with transplant centers outside their home state to benefit from multiple listings. This creates logistical and physical challenges as patients have to visit centers for monthly checkup.

[6]Transplant centers are not required to inform patients about these offers. In fact, the tight deadline (1 hour) makes it extremely difficult to include patients in the decision making process. As a result, transplant centers become the primary decision makers and only communicate accepted offers to the patients. Patients generally follow the surgeons recommendation(Husain et al., 2019). I present snapshots of the UNet system that is used to notify centers of any incoming kidney offers in the Appendix.

or July. Medicare examines the observed (O) and risk adjusted expected (E) 1-year graft failure or patient death based on a rolling 2.5-year transplant cohort.

The risk adjusted expected (E) 1-year graft failure or patient death is calculated using the Cox regression model. The model uses observations of all the patients and donors in the country and their characteristics and outcomes to estimate the effect of each characteristic on outcomes. The estimated effect are then applied to each patient-donor combination, giving an expected outcome for each patient, which will then be added up for all patients treated by the same transplant program. The list of variables used in risk adjustment models are reviewed and updated biannually by the SRTR analytics committee, getting opinions from surgeons around the U.S. I present examples of these variables in the Appendix. The rolling 2.5 year transplant cohort is best illustrated using the example in Figure 1. A report that is released on January 2008 is based on all transplants from July 2004 - January 2007 (i.e. months under the red line); a July 2008 report is based on all transplant from January 2005 - July 2007 (i.e. months above the grey line).

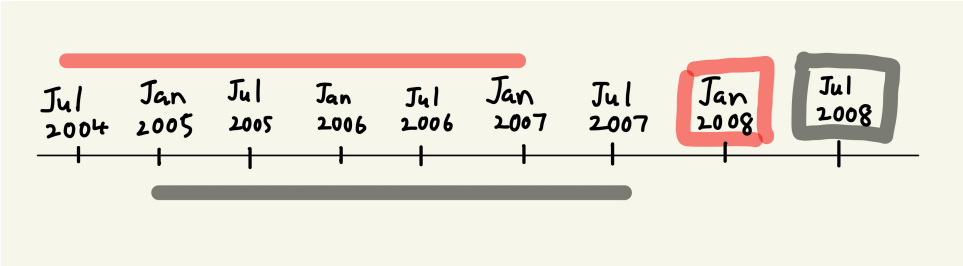


Figure 1: Examples of 2.5 year rolling cohort.

In order to be compliant, transplant centers must meet at least one of the following standards for 1-year graft failure or patient death:

1. Observed (O)/Expected (E) = OE ratio$\leq$ 1.5
2. p(Observed (O) - Expected (E) $\geq$ 0) = 1-sided p-value $\geq$ 0.05

The 1-sided p-value describes the probability that the observed difference is due to chance

alone. SRTR collects all the observed difference and compares it to all other transplant center in the U.S. accounting for the number of transplants, patients and donors managed by each transplant center. The 5% probability threshold highlights Medicare's comfort with the possiblity of misclassifying a center as under-performing. Figure 2 is an example of a CoP assessment for a transplant center. OE-ratio is calculated in line 8; 1-sided p-value is calculated by multiplying the value in line 10 by 1/2.

Line		Center 1 Year	National 1 Year
	Adult (Age 18+)		
1	Transplants (n=number)	90	10,781
2	**Percent (%) of Patients Surviving at End of Period**		
3	Observed at this Center	87.78	86.26
4	Expected, based on national experience	89.41	
5	**Deaths During Follow-up Period**		
6	Observed at this center	11	1,392
7	Expected, based on national experience	8.48	1,392
8	Ratio: Observed to Expected (O/E)	1.30	1.00
9	(95% Confidence Interval)	(0.65-2.32)	
10	P-value (2-sided), observed v. expected	0.469	
11	How does this center's survival compare to what is expected for similar patients?	Not Significantly Different (a)	
12	Percent retransplanted	5.5	4.4
13	Follow-up days reported by center (%)	91.7	93.9
14	Maximum Days of Follow-up (n)	365	365

Figure 2: Example of CoP assessment for a transplant center. The relevant information are found in lines 5 - 10. This is from (Dickinson et al., 2008).

Transplant centers who fail to meet all the conditions will be flagged and will be required by Medicare to maintain a data-driven quality assessment and performance improvement (QAPI) system. If the transplant center is flagged again within the next 30 months, it risks losing its program certification and funding from Medicare. These transplant centers have 210 days to appeal to Medicare that their outcomes are poor due to mitigating factors and how they intend to improve on performance.

3 Data

This paper uses the Standard Transplantation Analysis and Research (STAR) and Program Specific Report (PSR) dataset. The STAR dataset is available at the patient level and contains detailed information on all waitlisted patient and donor characteristics. Some patient charactheristics include age, BMI, diabetes status, date of waitlist and transplant and reasons for leaving waitlist. Some donor characteristics include age, reasons for death and kidney quality index. The PSR dataset is avaialbale at the transplant center level and are released biannually. They document transplant center activities 1 year prior to the release of the document. They also include the variables used for Medicare's CoP assessment as shown in Figure 2.

For the purpose of my study, I use observations from July 2007 to December 2012 [7]and follow the literature by dropping centers that specialize in pediatric transplant [8]. I aggregate patient-level information from the STAR dataset and combine it with PSR files to construct an unbalanced panel of transplant centers for every 6-months window (January to June or July to December). The 1st month of every 6-months window corresponds to the release of Medicare's CoP assessment. For example, January 2008 - June 2008 is a 6-months window and January 2008 is when the Medicare CoP report was released. My final dataset has 2278 center-window observations and covers 11 6-months window.

Table 1 describes my sample of centers. Approximately 8% - 12% of centers are flagged in each window. The "Flagged" column in Table 1specifies the number of flagged centers in the related 6-months window and is not cumulative[9]. Over the 11 windows, the number of centers have decreased with more exits than entries. Unfortunately, I do not observe the

[7]In 2014, a new allocation process termed "Longevity Matching" was introduced. The proposal was finalized in mid 2013 and implemented in 2014. Hence I decided to drop obeservations after 2012 to minimize any pre-policy interference.

[8]These centers contribute to about 2% of total transplant activity.

[9]For example, center A is flagged in 2007 07-12 and not 2008 01-06. Thus center A will enter 2007 07-12's flagged count and not 2008 01-06.

Windows	No. of Centers	Exit	Entry	Flagged
2007 07-12	210	0	0	19
2008 01-06	211	1	2	15
2008 07-12	210	2	1	24
2009 01-06	210	1	1	21
2009 07-12	210	1	1	24
2010 01-06	208	2	0	22
2010 07-12	207	2	1	19
2011 01-06	205	3	1	18
2011 07-12	205	0	0	15
2012 01-06	202	3	0	18
2012 07-12	200	2	0	13

Table 1: Number of centers between July 2007 and December 2012

exact reasons for centers closing down. It could be either poor performance or stakeholders finding the business not profitable.

No. of times centers are flagged	Frequency	Percent
0	153	70.51
1	16	7.37
> 2	48	22.12
Total	217	100

Table 2: Number of flagged incidence for each center between July 2007 and December 2012

Table 2 describes the number of times each unique transplant center was flagged during the sample period. 71% of centers were never flagged; 7% were flagged only once and 22% were flagged more than 2 times. Even though CoP threatens decertification after two flags in 30 months, these are rarely enforced immediately because most centers submit mitigating factors request and take advantage of the delay in enforcement to achieve some kind of performance improvement. Furthermore, centers are still allowed to operate when they undergo system improvements.

Table 3 compares characteristics of non-flagged and flagged center in my sample. Approximately 10% of observations are flagged. On average, non-flagged centers perform more

VARIABLES	(1) Non-Flagged mean	(2) sd	(3) Flagged mean	(4) sd
No. of Centers	2,070	-	208	-
Total TX Performed	27.87	23.98	18.29	15.22
– Low-Risk Kidneys	11.95	10.98	8.120	7.257
– Medium-Risk Kidneys	13.40	12.18	8.875	8.236
– High-Risk Kidneys	2.516	3.623	1.298	2.151
New Patients Recruited	86.29	79.99	50.89	43.83
– Age	50.33	13.94	49.76	14.18
– Body Mass Index (BMI)	28.21	5.734	28.10	5.710
– % Non-Diabetic Patient	0.448	0.497	0.430	0.495

Table 3: Summary Statistics of non-flagged and flagged centers across all 6-months windows

kidney transplants than flagged centers. This is also true if I divide the kidney offers into different kidney quality index [10]. Secondly, the average flagged center recruits less new patients than flagged centers in a 6-month window. The average patient in a flagged center is younger and less obese. However, the percentage of non-diabetic patient is lower in flagged centers.

4 Research Design

In this paper, I examine if transplant centers flagged by CoP for poor performance reduce (increase) the transplant of high (low) risk kidneys and waitlist patients associated with desirable characteristics (i.e. younger, less obese and non-diabetic). In particular, I estimate the ATE of being flagged for poor performance on the various outcome of interest, Y_{jt} (i.e. transplant volume for high or low risk kidney, average patient age, BMI and diabetic status). I take advantage of discontinuities in CoP's implementation (i.e. OE ratio > 1.5 and

[10] I follow Adler et al. (2014) and divide the kidneys according to the 3 different levels of kidney quality index. The most common kidney quality index is kidney donor profile index (KDPI). Source: https://optn.transplant.hrsa.gov/media/1512/guide-to-calculating-interpreting-kdpi.pdf

p-value < 0.05) and implement a sharp RD design. My parameter of interest is:

$$\beta = \mathbb{E}\left[Y_{jt}(1) - Y_{jt}(0) | (OE^r_{jt}, P_{jt}) = c\right] \tag{1}$$

β is the ATE at the cutoff $c = (1.5, 0.05)$. $\{Y_{jt}(1), Y_{jt}(0)\}$ are the usual Rubin Causal Model potential outcomes of transplant center j in window t; (OE^r_{jt}, P_{jt}) are the OE ratio and p-value of transplant center j in window t respectively. In the following subsections, I discuss the validity of the RD design in the kidney transplant setting and the process of estimating β.

4.1 Manipulation of Running Variables at the threshold

A key assumption of the RD design is that agents cannot precisely manipulate the running variable near the threshold (Lee and Lemieux, 2010). For simplicity, I examine the two running variables independently and focus most of my discussion on the oe-ratio. OE^r_{jt} denotes transplant center j's oe-ratio in window t.

A transplant center's OE^r_{jt} is based on all the transplants it performed during a 2.5 year rolling cohort. Variation in OE^r_{jt} arises from graft failure or patient death 1 year after the transplant. Centers control OE^r_{jt} in the sense that they have some authority in selecting the best patient-kidney combination that maximizes post-transplant outcome. However, there are also many idiosyncratic elements that determine a successful transplant. For example, patients with kidney failure often have significant commorbidities (Franczyk-Skóra et al., 2014). Furthermore, immunosuppressant drugs taken after the transplant increased the chance of disease infections. These two factors introduce health complications and reduce the likelihood of a successful transplant. It is in this sense that the transplant centers cannot precisely manipulate OE^r_{jt}.

Next, I use McCrary (2008)'s approach to formally test if there is manipulation of OE^r_{jt} at

the CoP threshold. The test uses local linear density estimator to estimate the discontinuity of the OE^r_{jt} density function near the threshold. If there is no manipulation of OE^r_{jt}, we should expect the density function to be smooth and continuous at the threshold. McCrary (2008) shows that under standard non-parametric regularity conditions, the discontinuity estimator is consistent and asymptotically normal. I refer readers interested in the details of the estimator to the Appendix.

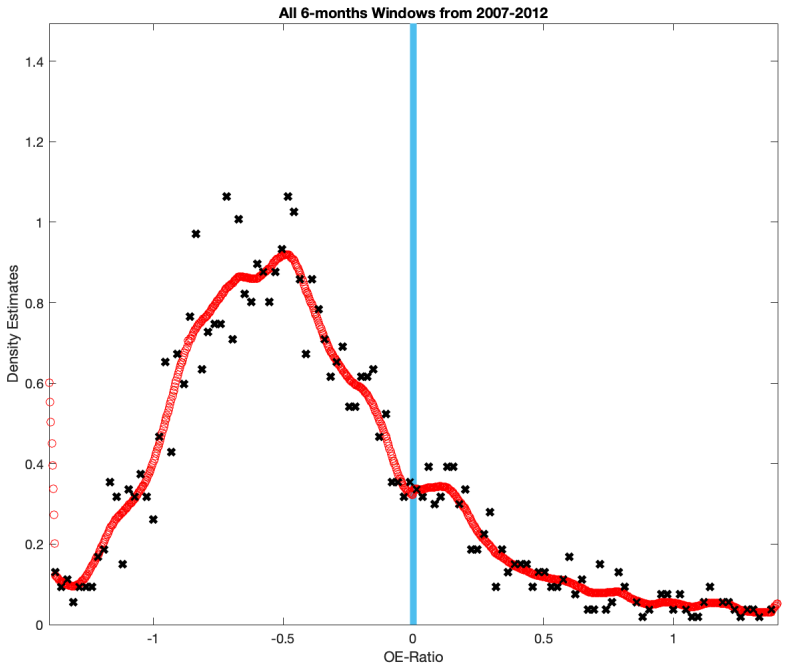


Figure 3: OE-ratio relative to the cutoff , Jul 2007 - Dec 2012

Figure 3 plots the estimate of density function along the support of OE^r_{jt}. I rescaled OE^r_{jt} such that the cutoff is now 0 as indicated by the blue line. The red circle outlines the local linear density estimator of point r, where r is a point along the support of OE^r_{jt}. Here we can see that there is no large or obvious discontinuity near the threshold. This is confirmed by the hypothesis test result in Column 1 Table 8 in the Appendix.

4.1.1 1st Step - Obtain Optimal Bandwidths (h_r^*, h_p^*)

Each jt observation in my panel data is a point on the (OE^r, P) grid, I use local linear regression analysis - with arbitrary bandwidths (h_r, h_p) - to estimate a fitted value of the outcome at all points (OE_{jt}^r, P_{jt})[11]:

$$\widehat{\mu}(OE_{jt}^r, P_{jt}, h_r, h_p) = \widehat{\gamma}_0 + \widehat{\gamma}_1 OE_{jt}^r + \widehat{\gamma}_2 P_{jt} + \widehat{\gamma}_3 (OE_{jt}^r \times P_{jt}^r) + \widehat{\delta_t} \tag{2}$$

δ_t is the 6-months window fixed effects. In each case, I limit the observations used to estimate $\widehat{\mu}(OE_{jt}^r, P_{jt}, h_r, h_p)$ as if it was a boundary point in order to mirror the regression-discontinuity approach that estimates limits defined at the boundary point of the region. For example, the subsample R_{jt} used to estimate $\widehat{\mu}(.)$ when $OE_{jt}^r \geq 0, P_{jt} \leq 0$ is:

$$R_{jt} = \{(Y_{ik}, OE_{ik}^r, P_{ik}) : (OE_{jt}^r \leq OE_{ik}^r < OE_{jt}^r + h_r) \cap (P_{jt} - h_p < P_k \leq P_{jt})\}$$

Subsample R_{jt} is depicted as the red square in Figure 4. Subsamples for other regions in the (OE^r, P) plane are defined analogously.

Next, I run OLS regression using observations in R_{jt} to obtain the fitted values $\widehat{\mu}(.)$ in equation 2. I then compare my fitted values to the observed values across the entire sample, using the generalized version of Imbens and Lemieux (2008)'s cross-validation criterion. The optimal bandwidths (h_r^*, h_p^*) are defined as the solution to the generalized cross-validation criterion:

$$\begin{aligned}(h_r^*, h_p^*) &= \arg\min_{h_r, h_p} CV^{IL}(h_r, h_p) \\ &= \arg\min_{h_r, h_p} \sum_j^N \sum_t^T (Y_{jt} - \widehat{\mu}(OE_{jt}^r, P_{jt}, h_r, h_p))^2\end{aligned} \tag{3}$$

[11] I follow Imbens and Lemieux (2008) and use the rectangular kernel to weight the different data points in the bandwidths in fitting the local linear regressions.

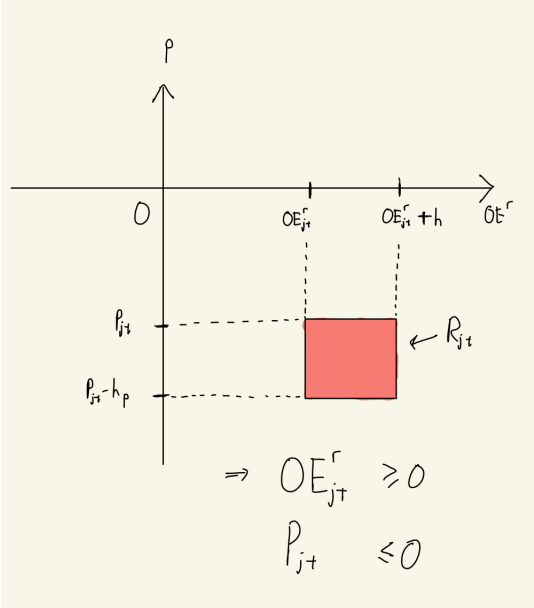


Figure 4: Region for local linear regression for each jt observation

4.1.2 2nd Step - Local Linear Regression with (h_r^*, h_p^*)

I use the optimal joint bandwidths (h_r^*, h_p^*) to identify the final local linear regression sample:

$$R^*(h_r^*, h_p^*) = \{(Y_{jt}, OE_{jt}^r, P_{jt}) : (|OE_{jt}^r| \leq h_r^*) \cap (|P_{jt}| \leq h_p^*)\} \tag{4}$$

and run an OLS regression of equation 5 using only observations in $R^*(h_r^*, h_p^*)$:

$$\begin{aligned} Y_{jt} = {} & \tau_0 + \beta F_{jt} + \tau_1 OE_{jt}^r + \tau_2 P_{jt} \\ & + \tau_3 (OE_{jt}^r \times P_{jt}) + \tau_4 (OE_{jt}^r \times F_{jt}) + \tau_5 (P_{jt} \times F_{jt}) \\ & + \tau_6 (OE_{jt}^r \times P_{jt} \times F_{jt}) + \delta_t + \epsilon_{jt} \end{aligned} \tag{5}$$

where F_{jt} is a dummy variable indicating if center j was flagged in window t; δ_t is the 6-months window fixed effects and ϵ_{jt} is center j idiosyncratic shock at window t. β is the parameter of interest as defined in equation 1. Next, I discuss result for different outcome of interest, Y_{jt}.

5 Results

5.1 Previous literature

I show that my data can replicate the results in Schold et al. (2013) and White et al. (2015). I follow the papers and run a simple linear regression regressing the outcome variables against a dummy variable of whether the transplant centers were flagged by CoP[12].

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
					Mean	Mean	Proportion of
VARIABLES	$\ln(TX_{jt})$	$\ln(TX_{jt}^{lr})$	$\ln(TX_{jt}^{mr})$	$\ln(TX_{jt}^{hr})$	Age_{jt}	BMI_{jt}	$NonDiab_{jt}$
Flagged, β	-0.3486***	-0.2737***	-0.3497***	-0.3015***	-0.9290***	-0.2704***	0.0228***
	[0.0750]	[0.0694]	[0.0703]	[0.0601]	[0.2555]	[0.0971]	[0.0083]
Observations	2,278	2,278	2,278	2,278	2,191	2,191	2,191
Window FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Results of previous literature

In columns 1-4 of Table 4, I am able to reproduce most of the results of Schold et al. (2013), flagged centers reduce kidney transplant activity except in column 2. In columns 5-7, I show that flagged centers waitlist older, more obese and less diabetic patients, similar to results found in White et al. (2015).

5.2 Transplant Activity

A hypothesis that has been proposed by the transplant community is that centers flagged for bad performance reduce the number of transplants performed (Schold et al. (2013)). The intuition is that transplant centers risk losing their funding or certfication if they are flagged again within the next 30 months. Thus flagged transplant centers become conservative to

[12]In reality, these papers did not run any formal regression and simply compare the means between treatment and control groups

reduce their exposure to unnecessary transplant failure.

I set $Y_{jt} = \ln(TX_{jt})$ in equation 5 and implement the estimation procedure described in section **??**. $\ln(TX_{jt})$ is the natural log of total number of kidneys transplanted by center j in period t. In column 1 of Table 5, the β estimate is not statistically significant, thus we cannot reject the null hypothesis that being flagged for bad performance does not affect a center's transplant activity.

Next, I test whether centers vary their transplant activity for different quality of kidneys[13]. I follow Adler et al. (2014) and divide all accepted kidney offers into three different quality index (i)Low Risk (LR) (ii) Medium Risk (MR) (iii) High Risk (HR). The intuition is that if flagged transplant centers are worried about their performance, we should expect them to transplant more low risk and less high risk kidneys. Thus the β estimates should be positive and negative respectively. This corresponds to columns 2 and 4 of Table 5. The β estimates display the expected signs, but are statistically insignificant.

I explore alternative variables that describe a center's transplant activity (i.e. rate of kidney transplant, TX_{jt}^{rate} and total kidney transplant, TX_{jt}) to check for the robustness of my results in Table 5. I present these results in Table 9 of Appendix E

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	$\ln(TX_{jt})$	$\ln(TX_{jt}^{lr})$	$\ln(TX_{jt}^{mr})$	$\ln(TX_{jt}^{hr})$	Mean Age_{jt}	Mean BMI_{jt}	Proportion of $NonDiab_{jt}$
Flagged, β	0.10897	0.87129	0.0055208	-0.63528	0.39411	1.0259	-0.048131
	[0.6152]	[0.64197]	[0.59731]	[0.7055]	[3.1346]	[1.0495]	[0.089669]
Observations	59	62	59	62	67	67	67
Window FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Effect of CoP on center transplant or waitlist strategy (nonparametric RDD)

[13]The most common kidney quality index is kidney donor profile index (KDPI). Source: https://optn.transplant.hrsa.gov/media/1512/guide-to-calculating-interpreting-kdpi.pdf

5.3 Selective Waitlisting

The issue of selective waitlisting has been discussed in the transplant community (White et al., 2015) and there are anectdotal evidence to suggest its existence [14]. Here, I use CoP to study selective waitlisting. The intuition is that: CoP incentivizes centers to address performance concerns by rejecting unhealthy patients. I test this with equation 5, replacing Y_{jt} with some characteristics that are anticipated to be undesirable (i.e. old age; obese and diabetic patients).

I took patient-level information on age, BMI and diabetes status and averaged them to center-level. Y_{jt} is now replaced with mean age, mean BMI and proportion of non-diabetic patient. The β estimates across the columns 5-7 in Table 5 are not statistically significant, thus we cannot reject the null hypothesis that being flagged for bad performance does not affect a center's waitlist strategy.

5.4 Discussion and Robustness Check

In this subsection, I discuss my decision to use local linear regression instead of relying on a parametric approach. Gelman and Imbens (2019) showed that using high-order (≥ 3) polynomials in RD analysis leads to noisy estimates and poor coverage of confidence intervals. They instead recommend researchers to use estimators based on quadractic polynomials or local linear regression.

I follow Gelman and Imbens (2019) and assume a linear and quadratic regression function. I present the results in Table 6 and 7 respectively[15]. Results in Column 1 to 4 of both tables are different from the corresponding columns in Table 5, suggesting that flagged centers lower their transplant activity after being flagged for poor performance. Results in Column

[14]Source: https://www.statnews.com/2016/08/11/organ-transplant-federal-standards/

[15]Columns 1-4 have 2278 observations and columns 5-7 have 2191 observations. This difference is because some transplant center did not recruit any patient during the 6-month window.

5 - 7 of both tables are still broadly consistent with the correspondings columns in Table 5. The discrepancies suggest that my results are sensitive to the particular approach I use in my estimation, but I want to argue that the local linear regression is more appropriate in my setting.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	$\ln(TX_{jt})$	$\ln(TX^{lr}_{jt})$	$\ln(TX^{mr}_{jt})$	$\ln(TX^{hr}_{jt})$	Mean Age_{jt}	Mean BMI_{jt}	Proportion of $NonDiab_{jt}$
Flagged, β	-0.9589***	-0.8154***	-0.9410***	-0.5640***	0.0842	-0.3037	0.0508**
	[0.2344]	[0.2169]	[0.2191]	[0.1898]	[0.7908]	[0.2977]	[0.0257]
Observations	2278	2278	2278	2278	2191	2191	2191
Window FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Effect of CoP on center behavior (linear $\mathbb{E}[Y|X]$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	$\ln(TX_{jt})$	$\ln(TX^{lr}_{jt})$	$\ln(TX^{mr}_{jt})$	$\ln(TX^{hr}_{jt})$	Mean Age_{jt}	Mean BMI_{jt}	Proportion of $NonDiab_{jt}$
Flagged, β	-0.5982**	-0.7004**	-0.4830*	-0.3817	1.1560	0.1446	-0.0381
	[0.2914]	[0.2786]	[0.2769]	[0.2560]	[1.1172]	[0.4203]	[0.0363]
Observations	2,278	2,278	2,278	2,278	2,191	2,191	2,191
Window FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

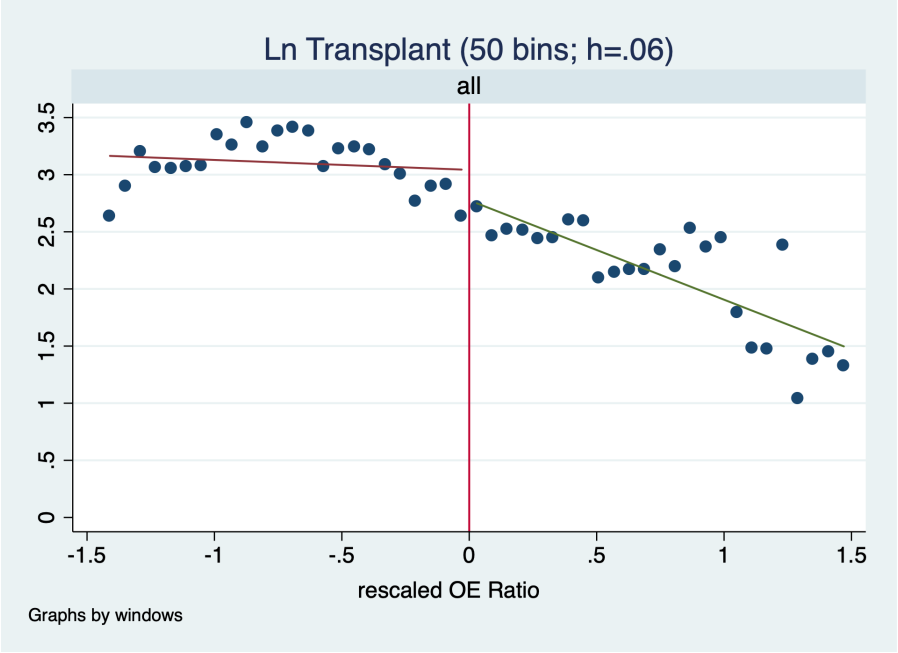
Standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Effect of CoP on center behavior (quadratic $\mathbb{E}[Y|X]$)

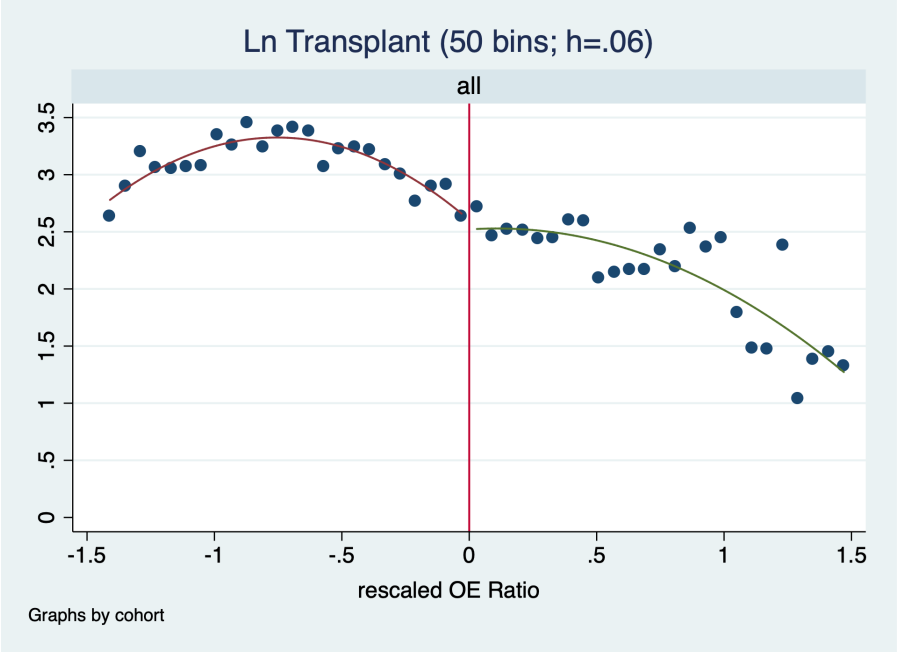
For simplicity, I follow Lee and Lemieux (2010)'s recommendation and present a bin-scatter plot of $\ln(TX_{jt})$ against OE^r_{jt}. I provide evidence in Figure 5 that suggest both linear and quadratic polynomials might not be a good fit for the conditional expectation function. This pattern is also true for the other outcome of interest, Y_{jt} (i.e. mean age or BMI of recruited patients and transplant of high or low risk kidneys).

The advantage of the nonparametric approach is that it only uses points that are close to the cutoff. This is in line with the spirit of RD design that depends on local estimates of the regression function at the cutoff. The disadvantage is that my results in Table 5 only

use approximately 3% ($\frac{60}{2278}$) of my total sample and lack external validity, in the sense that I cannot comment on center behavior far away from the cutoff.



(a) Linear Regression Function



(b) Quadratic Regression Function

Figure 5: Linear and quadratic regression functions might not be good fit.

6 Conclusion

I use a sharp RD design to estimate the ATE of being flagged by CoP for poor performance on transplant center incentives. Contrary to previous literature, I find no significant results to suggest that flagged transplant centers change their transplant or waitlist strategy. Transplant volume of both high and low risk kidneys remain unchanged; centers do not waitlist younger, less obese or more non-diabetic patients.

My work have serious limitations. For example, the RD estimates from local linear regression have internal validity but lack external validity. It does not allow me to make any serious statement about the behavior of transplant centers far away from the cutoff. Furthermore. my work only deals with aggregate data at transplant center level. It examines a very specific channel and assumes that centers only change their behavior after being flagged. This might not reflect reality.

In an ongoing analysis, I examine other channels where CoP affect transplant center incentives. In particular, I explore the possibility that centers have information or tools that allow them to track changes in OE^r scores in real time and change their strategies dynamically. The intuition is: if OE^r is approaching the cutoff (1.5), transplant centers become more conservative (i.e. avoid high-risk kidneys or only accept healtheir patients). Likewise, if OE^r is below and very far from 1.5, they are not too concerned with their performance even if they were already flagged. If the variation in OE^r is exogenous, I can exploit within-center covariation between OE^r and transplanted kidney/waitlisted patient characteristics to identify changes in center behavior. This alternative mechanism is probably more realistic and allows me to take advantage of my patient-level dataset.

Moving forward, there are also other interesting research question. How should we design the different features of CoP? For example: Is the current 6-month reporting period too long or too short? Is the OE ratio 1.5 threshold appropriate? How will the modifications affect

the different sizes of transplant centers? To answer these questions, I will need to build a structural model to conduct counterfactual analysis.

References

Adler, Joel T., Rosh K. V. Sethi, Heidi Yeh, James F. Markmann, and Louis L. Nguyen. 2014. "Market Competition Influences Renal Transplantation Risk and Outcomes." *Annals of Surgery*, 260(3): 550–557.

Agarwal, Nikhil, Itai Ashlagi, Michael A. Rees, Paulo Somaini, and Daniel Waldinger. 2021. "Equilibrium Allocations Under Alternative Waitlist Designs: Evidence From Deceased Donor Kidneys." *Econometrica*, 89(1): 37–76.

American Kidney Fund (AKF). 2008. "Deceased donor kidney transplants."

Bundorf, M. Kate, Natalie Chun, Gopi Shah Goda, and Daniel P. Kessler. 2009. "Do markets respond to quality information? The case of fertility clinics." *Journal of Health Economics*, 28(3): 718–727.

Dickinson, D. M., C. J. Arrington, G. Fant, G. N. Levine, D. E. Schaubel, T. L. Pruett, M. S. Roberts, and R. A. Wolfe. 2008. "SRTR program-specific reports on outcomes: A guide for the new reader." *American Journal of Transplantation*, 8(4 PART 2): 1012–1026.

Dranove, David, and Ginger Zhe Jin. 2010. "Quality Disclosure and Certification: Theory and Practice." *Journal of Economic Literature*, 48(4): 935–63.

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

Franczyk-Skóra, Beata, Anna Gluba, Robert Olszewski, Maciej Banach, and Jacek Rysz. 2014. "Heart function disturbances in chronic kidney disease – echocardiographic indices." *Archives of Medical Science*, 6: 1109–1116.

Gelman, Andrew, and Guido Imbens. 2019. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics*, 37(3): 447–456.

Hamilton, Thomas. 2013. "Regulatory Oversight in Transplantation: Are the Patients Really Better Off?" *Current opinion in organ transplantation*, 18.

Held, P.J., F. McCormick, A. Ojo, and J.P. Roberts. 2016. "A Cost-Benefit Analysis of Government Compensation of Kidney Donors." *American Journal of Transplantation*, 16(3): 877–885.

Husain, S. Ali, Kristen L. King, Stephen Pastan, Rachel E. Patzer, David J. Cohen, Jai Radhakrishnan, and Sumit Mohan. 2019. "Association Between Declined Offers of Deceased Donor Kidney Allograft and Outcomes in Kidney Transplant Candidates." *JAMA Network Open*, 2(8): e1910312–e1910312.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression discontinuity designs: A guide to practice." *Journal of Econometrics*, 142(2): 615–635. The regression discontinuity design: Theory and applications.

Kolstad, Jonathan T. 2013. "Information and Quality When Motivation Is Intrinsic: Evidence from Surgeon Report Cards." *American Economic Review*, 103(7): 2875–2910.

Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature*, 48(2): 281–355.

Locke, Jayme E. 2018. "Handbook of Kidney Transplantation. Sixth Edition. G. M. Danovitch (Editor) Wolters Kluwar, 2017, 606pp. ISBN: 9781496326157." *American Journal of Transplantation*, 18(7): 1820–1820.

McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics*, 142(2): 698–714.

Ramanarayanan, Subramaniam. 2011. "Reputations and Firm Performance: Evidence from the Dialysis Industry."

Schold, J.D., L.D. Buccini, T.R. Srinivas, R.T. Srinivas, E.D. Poggio, S.M. Flechner, C. Soria, D.L. Segev, J. Fung, and D.A. Goldfarb. 2013. "The Association of Center Performance Evaluations and Kidney Transplant Volume in the United States." *American Journal of Transplantation*, 13(1): 67–75.

Schold, Jesse. 2020. "The evolving role of regulatory reporting on patient and donor selection in organ transplantation." *Current Opinion in Organ Transplantation*, 25: 1.

Stith, Sarah S., and Richard A. Hirth. 2016. "The Effect of Performance Standards on Health Care Provider Behavior: Evidence from Kidney Transplantation." *Journal of Economics & Management Strategy*, 25(4): 789–825.

White, Sarah L., Dawn M. Zinsser, Matthew Paul, Gregory N. Levine, Tempie Shearon, Valarie B. Ashby, John C. Magee, Yi Li, and Alan B. Leichtman. 2015. "Patient Selection and Volume in the Era Surrounding Implementation of Medicare Conditions of Participation for Transplant Programs." *Health Services Research*, 50(2): 330–350.

Wolfe, R. A., K. P. McCullough, D. E. Schaubel, J. D. Kalbfleisch, S. Murray, M. D. Stegall, and A. B. Leichtman. 2008. "Calculating Life Years from Transplant (LYFT): Methods for Kidney and Kidney-Pancreas Candidates." *American Journal of Transplantation*, 8(4p2): 997–1011.

Zhang, Juanjuan. 2010. "The Sound of Silence: Observational Learning in the U.S. Kidney Market." *Marketing Science*, 29(2): 315–335.

Appendices

A Biology of Kidney Transplant

A kidney from a decased donor is considered transplantable to a patient if they are biologically comptaible. A donor's organ is considered incompatible if the patient has a pre-existing immune response to proteins on the organ's cells. A biologically incompatible patient's immune system will recognize and attack the transplanted organ, resulting in graft failure or patient death. Following transplantation, medications allow transplant physicians to limit new immune responses to foreign protein types. Interested readers can refer to Locke (2018) for further details on kidney biology.

B Geography of Kidney Market

Organ Procurement Organizations (OPOs) are not-for-profit organizations responsible for recovering organs from deceased donors for transplantation in the U.S. There are 58 OPOs assigned to their individual donor service area (DSA). The OPO's role is to assess donor potential, collect and convey accurate clinical information and follow national policies for offering organs. For every successful match, the OPO facilitates authorization, testing, the recovery of donor organs and delivery to the transplant center[16].

[16]Source: https://unos.org/transplant/opos-increasing-organ-donation/

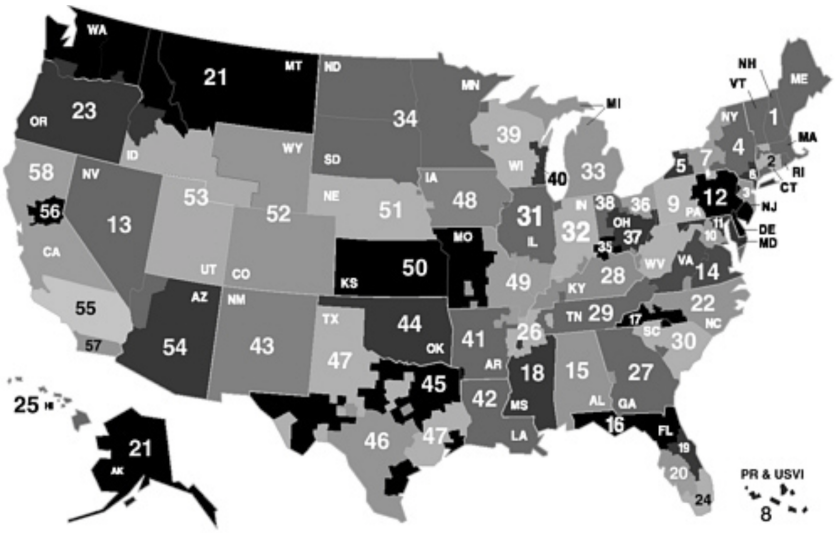


Figure 6: 58 OPO and their DSA

C UNet System

UNet is an online platform that transplant center view all their incoming kidney offers. For example, when a kidney offer arrives, New York Columbia Hospital (NYCH) observes information on the kidney:

Seq	Donor ID	Age	Blood	BMI	Cause of death	Date & Time of death
1	A88	36	O	24.3	Car Accident	August 8th 2020 1600 ET

Table: Characteristics of Kidney Offer

NYCH can observe all the compatible patients. But they will not know the identity or characteristics of the patients not in NYCH.

Seq	Center	Name	Age	Blood	BMI	Dialysis Time	Offer Respond (Y/N)
1	NYCH	Bob	56	O	22.5	5.6 years	
2	***	***	45	B	***	***	N
3	***	***	68	O	***	***	
4	NYCH	Charlie	36	O	30	2.3 years	
5	***	***	39	AB	***	***	Y

Table: Characteristics of Compatible Patients

D Manipulation of OE^r_{jt}

Here I describe the test by McCrary (2008) that formally estimates the discontinuity in the density function of the running variable at the threshold. The first step histogram is based on the frequency table of a discretized version of the running variable:

$$g(OE^r_{jt}) = \left\lfloor \frac{OE^r_{jt} - c}{b} \right\rfloor b + \frac{b}{2} + c \in \left\{ .., c - \frac{3b}{2}, c - \frac{b}{2}, c + \frac{b}{2}, c + \frac{3b}{2} ... \right\} \tag{6}$$

where $\lfloor a \rfloor$ is the greatest integer in a. c is the cutoff value for OE^r_{jt} and b is the width covering the support of $g(OE^r_{jt})$. Next, I define an equi-spaced grid $X_1, X_2, ..., X_M$ of width b covering the support of $g(OE^r_{jt})$ and the normalized cellsize for the mth bin:

$$Y_m = \frac{1}{Nb} \sum_{j=1}^{J} \sum_{t=1}^{T} \mathbb{1}(g(OE^r_{jt} = X_m)) \tag{7}$$

The second step smooths the histogram using local linear regression. The density estimate at r is give by $\widehat{f}(r) = \widehat{\phi}_1$, where:

$$(\widehat{\phi_1}, \widehat{\phi_2}) = \arg\min_{\phi_1, \phi_2} \sum_{m=1}^{M} \{Y_m - \phi_1 - \phi_2(X_m - r)\}^2 K\left(\frac{X_m - r}{h}\right) \{\mathbb{1}(r \geq c \cap X_m > c) + \mathbb{1}(r < c \cap X_m < c)\} \tag{8}$$

where $K(t) = \max\{0, 1 - |t|\}$ and h is the bandwidth.The second step smooths the histogram by estimating a weighted regression using the bin midpoints to explain the height of the bins, giving most weight to the bins nearest to the point r.

Define the parameter of interest to be the log difference in height:

$$\theta = \ln \lim_{r \downarrow c} f(r) - \ln \lim_{r \uparrow c} f(r) = \ln f^+ - \ln f^- \tag{9}$$

Both f^+ and f^- can be obtained by estimating two separate local linear regression, on either side of c. We have:

$$\begin{aligned}\widehat{\theta} &= \ln \widehat{f^+} - \ln \widehat{f^-} \\ &= \ln \left\{ \sum_{X_m > c} K\left(\frac{X_m - c}{h}\right) \frac{S^+_{n,2} - S^+_{n,1}(X_m - c)}{S^+_{n,2} S^+_{n,0} - (S^+_{n,1})^2} Y_m \right\} \\ &- \ln \left\{ \sum_{X_m < c} K\left(\frac{X_m - c}{h}\right) \frac{S^-_{n,2} - S^-_{n,1}(X_m - c)}{S^-_{n,2} S^-_{n,0} - (S^-_{n,1})^2} Y_m \right\}\end{aligned}$$

where $S^+_{n,d} = \sum_{X_m > c} K\left(\frac{X_m - c}{h}\right)(X_m - c)^d$ and $S^-_{n,d} = \sum_{X_m < c} K\left(\frac{X_m - c}{h}\right)(X_m - c)^d$. In McCrary (2008), the paper shows that under standard non-parametric regularity conditions, $\widehat{\theta}$ is consistent and asymptotically normal:

$$\sqrt{nh}(\widehat{\theta} - \theta) \xrightarrow{d} N\left(B, \frac{24}{5}\left(\frac{1}{f^+} + \frac{1}{f^-}\right)\right) \quad \text{where} \quad B = \frac{H}{20}\left(\frac{-f^{+"}}{f^+} - \frac{-f^{-"}}{f^-}\right)$$

and the approximate standard error for $\widehat{\theta}$ is:

$$\widehat{\sigma_\theta} = \sqrt{\frac{24}{5Nh}\left(\frac{1}{f^+} + \frac{1}{f^-}\right)}$$

	OE^r_{jt}
$\widehat{\theta}$	0.0433
	(0.3268)
Observations	2218

Table 8: Log discontinuity estimates

E Additional Results

VARIABLES	(1) TX_{jt}^{rate}	(2) $(TX_{jt}^{rate})^{lr}$	(3) $(TX_{jt}^{rate})^{mr}$	(4) $(TX_{jt}^{rate})^{hr}$	(5) TX_{jt}	(6) TX_{jt}^{lr}	(7) TX_{jt}^{mr}	(8) TX_{jt}^{hr}
Flagged, β	-0.051268	0.0005224	-0.0814847	-0.049027	2.4917	3.0339	2.5768	-1.8812
	[0.09043]	[0.11918]	[0.11222]	[0.070156]	[12.59]	[6.5564]	[0.37513]	[-0.81711]
Observations	62	62	62	62	59	62	59	61
Window FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table 9: Effect of CoP on transplant center activity